Exploring the world of non-coding genes in stem cells and autoimmunity.

Messemaker, T.C.

Exploring the world of non-coding genes in stem cells and autoimmunity

Tobias Casper Messemaker

Cover design by Hennie Messemaker
Printed by Ipskamp printing
ISBN: 978-94-028-0959-6

The work presented in this thesis was supported by the Dutch Arthritis Foundation (Reumafonds). Printing of this thesis was financially supported by the Leiden University.

Exploring the world of non-coding genes in stem cells and autoimmunity

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 3 april 2018
klokke 10.00 uur

door

Tobias Casper Messemaker
geboren te Katwijk aan Zee
in 1988

Promotor:	Prof. dr. R.E.M. Toes
Co-Promotores:	Dr. B.A.S. Kurreeman Dr. H.M.M. Mikkers
Leden promotiecommissie:	Prof. dr. T.W.J. Huizinga Prof. dr. R.C. Hoeben Prof. dr. J.H. Gribnau (Erasmusc MC) Prof. dr. A.P. Zhernakova (UMC Groningen)

The research presented in this thesis was performed at the department of Molecular Cell Biology and the department of rheumatology, Leiden University Medical Center, Leiden, The Netherlands.

Table of contents

Chapter 1

General introduction

Autoimmune diseases (AIDs) are common and can affect a wide variety of organs. Understanding why the immune system attacks the body's own cells is crucial in order to treat patients and prevent the onset of autoimmunity. In the past 100 years great efforts have been undertaken to gather insight into AIDs. Despite the advances, these studies have revealed that the complexity of AIDs is enormous. A wide range of AIDs exists. A few of the most common AIDs are rheumatoid arthritis, systemic lupus erythematosus, type I diabetes, thyroiditis, multiple sclerosis and psoriasis[1–3]. However, not all AIDs affect a large proportion of the population. For example the prevalence of systemic sclerosis is ~100 times lower than rheumatoid arthritis[4]. AIDs display a typical preference for females albeit the reason for this is still unknown[5–8]. Both genetic and environmental factors contribute to dysregulation of the immune system and disease pathogenesis. Some of these genetic and environmental factors overlap among different AIDs, but also disease-specific factors have been identified[9,10]. Genes identified through genetic studies, have pinpointed to the involvement of multiple pathways, which also act in cell-type specific manners[10,11]. Multiple environmental factors have been identified and are thought to play a role in the onset and development of AIDs. These include smoking, exposure to UV, microbes, nutrients and exposure to organic substances[12–15]. This variety of contributing factors illustrates the complexity of AID and indicates why causal factors are notoriously hard to be identified. In this thesis, rheumatoid arthritis and systemic sclerosis were more closely investigated. Which genes play a role and how these genes are deregulated were the main objectives of these studies. Moreover, the role of non-coding RNAs was studied in the context of rheumatoid arthritis, systemic sclerosis, and more basic transcriptional regulation.

Rheumatoid arthritis

Rheumatoid arthritis (RA) is the most common autoimmune disease with a prevalence of 0.5 to 1% in the adult population worldwide[16]. Prime characteristics of RA are inflammation of the joints leading to cartilage damage and bone damage. Many cell types are involved in this process including T-cells, B-cells, monocytes, macrophages, dendritic cells, synovial fibroblasts, synoviocytes, neutrophils, osteoclasts and mast cells[17]. These cell types are involved in i.) recognizing the self-proteins as foreign proteins, ii.) enhancing inflammation by cytokine production and the recruitment of other immune cells and iii.) secreting

enzymes involved in bone erosion and destruction[18]. The interplay between these cells and processes likely contributes to a self-stimulating process that results in the chronic nature of RA. The disease is more prevalent in women with a 2-3x higher incidence[19]. RA is a heterogeneous disease indicated by both seropositive and seronegative patients. Seropositivity is indicated by various autoantibodies, and associates with more severe symptoms, joint damage and higher mortality[20–24]. The most prevalent autoantibody known is Rheumatoid Factor (RF), which recognizes the Fc part of an IgG molecule. RF is found in approximately 75% of patients, however this autoantibody is also found in other diseases and in healthy individuals upon ageing[25]. A more RA-specific autoantibody is the anti-citrullinated antibody (ACPA), which is directed against citrullinated proteins. ACPAs are found in approximately 70% of patients and are highly specific for RA[26]. A more recently discovered autoantibody in RA patients is the anti-carbamylated protein antibody (anti-CarP), which recognizes carbamylated proteins. These anti-CarP antibodies are present in ~40% of the patients and associate with disease activity and bone damage[23,27]. The positivity for some of these autoantibodies is linked to environmental factors. ACPA-positivity and RF-positivity are both higher in patients who have been smoking compared to non-smokers, while no specific relationship with anti-Carp positivity exists[27,28]. Interestingly these autoantibodies are present years before disease onset and can therefore be used for diagnosing RA[29,30]. Although these autoantibodies are key in diagnoses, classification, and prediction of disease severity, they are currently not exploited as targets for treatment.

Currently, the best treatment is provided by inhibition of pro-inflammatory cytokines. In RA patients, cytokines like TNFα and IL6 are elevated in the serum and synovium compared with healthy individuals[31–33]. Both anti-TNFα and anti-IL6 treatment are currently successful in alleviating rheumatoid arthritis[34,35]. TNFα is a potent inducer of inflammatory genes, resulting in increased local inflammation and bone degradation[33,34]. Prominent TNF inhibitors include infliximab, etanercept, adalimumab, golimumab and certolizumab pegol[36]. IL6 is a cytokine that activates the immune response of several cell types and is also involved in the maturation of B-cells[37]. Tocilizumab is an IL6-inhibitor which can bind soluble and membrane-bound IL6-receptors and is used as a therapeutic strategy to treat RA[38]. However, the mechanism by which the immune system is activated and the mechanism by which TNFα and IL6 are enhanced remain elucidative.

Both environmental and genetic components have been identified in RA. On basis of twin studies, the genetic component is estimated to account for 60% of the susceptibility to RA and is even more contributing in seropositive RA[39,40]. Besides familial or twin studies, a genetic contribution can also be investigated by analysing frequencies of genetic variants in large populations referred to as genome wide association studies (GWAS). In GWAS, the prevalence of genetic variants (such as single nucleotide polymorphism (SNPs)) in a disease population is compared with the prevalence of the same genetic variants in a healthy population thereby investigating whether certain variants are more frequent in diseased individuals. GWAS have identified over 100 associated loci that contribute to RA[41]. The strongest associating variants are located in the *HLA*-region on chromosome 6, which explains approximately 80% of the genetic contribution[41]. Specifically, variants in the *HLA-DRB1* gene, a gene involved in peptide presentation, associates strongly with RA susceptibility. The RA associated variants encode amino acid sequences in the peptide-binding groove, which is known as the shared epitope (SE)[42]. The SE epitope includes QKRAA, QQRAA and KKRAA on position 70-74 of the *HLA-DRB1* chain and points to a crucial role for peptide (and self-peptide) binding in RA pathogenesis[43].

Next to the association of the HLA locus, many non-HLA regions have been associated to RA. These non-HLA regions have lower odds ratios and are probably involved with a smaller functional contribution[41]. Several studies have investigated how the variants in non-HLA genes may translate to the increased onset and development of RA and other AIDs[44]. For example, *PTPN22* is a gene which acts as a negative regulator of the T-cell receptor of which several variants have been associated with multiple autoimmune diseases[45]. These variants are thought to interfere with PTPN22 functioning resulting in a diminished inhibitory effect and therefore increased T-cell activation[46,47]. Overall, genes located in these non-HLA regions are significantly enriched for immune-related pathways like NF-kb signalling pathway, T-cell receptor signalling pathway and the JAK-STAT signalling pathway[44]. Likely, genetic variants in the non-HLA regions disrupt genes within these pathways making an individual more susceptible to inflammatory diseases like RA. Together, RA is a multifactorial disease influenced by genetic and environmental factors, which together likely result in disturbed immune homeostasis in synovial areas eventually leading to disease pathogenesis, figure 1.

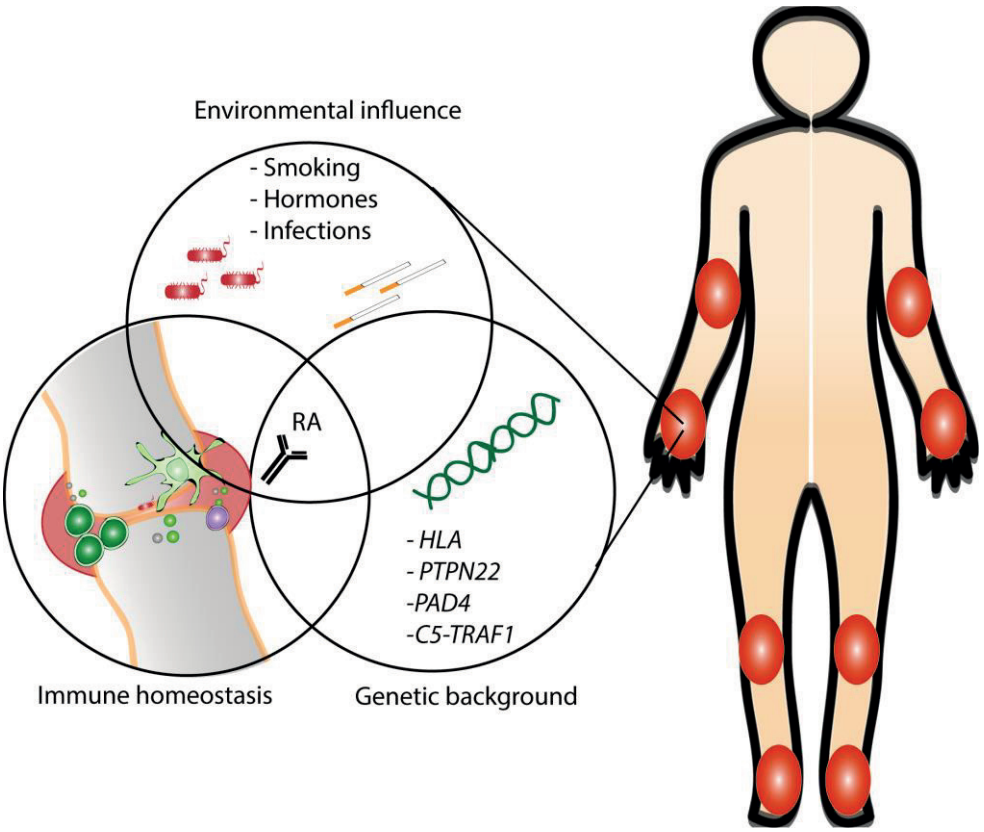

Figure 1. Schematic overview of factors influencing rheumatoid arthritis.

Systemic sclerosis

Systemic sclerosis (SSc) is a heterogeneous autoimmune disease with a prevalence of ~0.02% in the western population[4]. Similar to rheumatoid arthritis, females are more prone to develop this autoimmune disease with an observed female-to-male ratio of up to 1:5[48,49]. The typical diagnosis of SSc patients is based on fibrosis of the skin and complications of other internal organs. Over 90% of patients show skin fibrosis, ~90% gastrointestinal complications, ~65% musculoskeletal problems, ~40% interstitial lung disease, and ~15% of patients suffer from pulmonary arterial hypertension (PAH)[50]. Moreover typical characteristics of SSc are vascular manifestations reminiscent of Raynaud phenomenon, which are often observed prior to diagnosis of SSc[51]. SSc patients are grouped into limited cutaneous systemic sclerosis (lcSSc) and diffuse

cutaneous systemic sclerosis (dcSSc) on basis of their skin involvement. In lcSSc, skin involvement is restricted to the region between fingers and elbow, and face, while in dcSSc proximal regions are also affected[52]. Overall, dcSSc patients display a rapid disease progression with extensive skin fibrosis and development of complications of the internal organs but severity of the disease may differ between patients and between disease-subtypes[52].

A number of autoantibodies have been detected in SSc. The presence of autoantibodies is used for SSc diagnosis and classification of patients. Autoantibodies are mainly directed against nuclear components and are described as anti-nuclear antibodies (ANAs). ANAs include anti-centromeric antibodies, anti-topoisomerase I, anti-RNA polymerase III, anti-U1-RNP, anti-U3-RNP, anti-Th/To, anti-Pm/Scl and anti-nucleolar antibodies[53]. lcSSc patients display a stronger association with anti-centromeric antibodies, while dcSSc patients often have anti-topoisomerase and anti-RNA polymerase antibodies[53]. Finally, the complexity and heterogeneity of SSc is illustrated by the fact that some individuals are positive for SSc serology but lack the presence of detectable skin involvement[54].

Similar to RA and other AIDs, environmental and genetic components have been identified in SSc, figure 2. Known environmental factors are pollutants and chemicals including silica dust, vinyl chloride and organic substances[55,56]. Moreover, infectious agents, like viruses, have been reported to be associated with risk of developing SSc[55]. Despite numerous studies, the molecular mechanisms underlying SSc remain elusive. Approximately 40 genes have been linked to SSC by multiple genetic studies[57–59]. The *HLA*-locus (*HLA-DR* and *HLA-DP*) shows the strongest genetic association with SSc, indicating that (self) antigen presentation plays a role in SSc. Clinical implications of these genetic associations have been shown by correlation analysis with the presence of autoantibodies. Anti-topoisomerase antibodies correlated strongly with DPB1*1301 and DRB1*1101–21, while anti-centromeric antibodies were positively correlated with the presence of DRB1*0401–22 and DRB1*0801–11[60]. Besides the *HLA*-genes, other immunological genes are enriched in SSc associated regions, for example genes belonging to the interferon pathway[57].

Besides genetic evidence expression studies have shown that interferon genes are deregulated in patients. Expression studies have been performed in SSc tissues to investigate deregulated genes and altered pathways, and to discover

new drug targets. Affected tissues that have been investigated included skin, cultured fibroblasts, keratinocytes and various cells of the immune system. Besides, gene expression profiles are under investigation for the classification of SSc patients[61–64]. Milano *et al*. showed that patients can be subdivided into four groups on basis of gene expression profiles in the skin: i) deregulated expression of proliferative genes, ii) altered expression of inflammatory genes, iii) aberrant expression of fibrotic genes or iv) a normal-like gene expression profile[64].

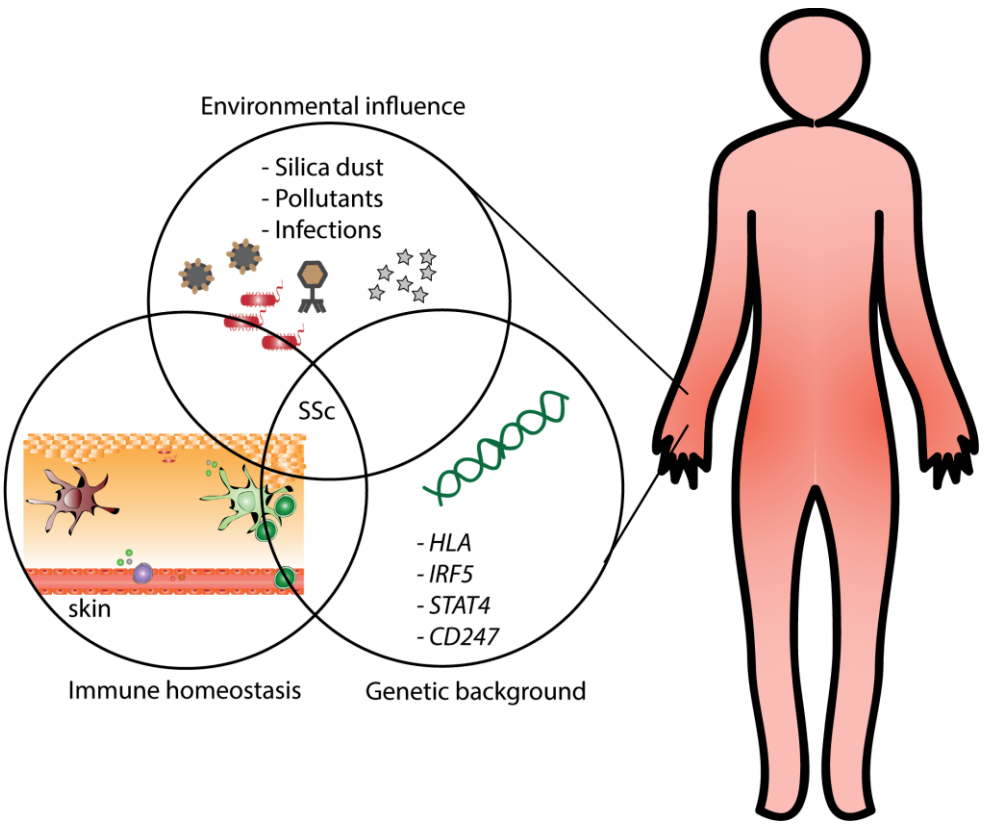

Figure 2. Schematic overview of factors influencing systemic sclerosis.

Although other studies did not classify SSc patients they observed deregulation of genes of immunological and fibro-proliferative nature[62,63]. Assassi *et al*. showed that many keratin-related genes are altered in SSc patients and that this keratin signature associate with early patients. On the other hand, a fibroinflammatory

signature correlates with dcSSc and higher skin scores[62]. Studies investigating gene expression in immune cells of SSc patients have described increased expression of macrophage and interferon genes including *CCR1*, *IL1B*, *IL13* and *JAK2*[65,66]. Together, these studies point to a disease mechanism in which a genetically primed individual is exposed to environmental factors triggering a chronic inflammatory process. This inflammatory process is characterized by vascular alterations and fibrosis. Early responses are likely mediated through macrophages, which induce the expression of interferon-related cytokines along with fibrotic mediators such as TGFβ. A continuous interplay between these pathways further exacerbates the fibrotic phenotype of the skin of patients which might be propelled further into other internal organs[67].
Current treatment of SSc is limited and consists mainly of immunosuppressive medication and haematopoietic stem cell transplantation. However, anti-fibrotic compounds are under investigation[68,69]. While current drug targets are based on coding genes only, a large proportion of the transcriptome is annotated as non-coding[70]. Thus investigating deregulated non-coding genes in SSc patients may lead to the identification of novel SSc biomarker genes and potentially new druggable targets.

Gene transcription and regulation and its relevance to rheumatic diseases.
Gene transcription and regulation is fundamental for all biological processes. Disturbed regulation of gene expression may lead to disease, including autoimmunity. Thus, insight into the complexity of gene regulation has broad implications for disease understanding and disease treatment.
In the early 1960's Francois Jacob and Jacques Monod pioneered the first model for gene regulation and introduced the phenomenon of gene activation and repression[71]. In the years after, many additional factors have been described to influence this process. Recruitment of RNA polymerases is essential for a gene to be transcribed into RNA. Recruitment and activation of RNA polymerases is mediated by transcription factors (TFs), which can bind specific DNA sequences known as promotors and enhancers. The interaction between DNA and transcription factors is defined by both the DNA sequence and the accessibility of the DNA. Importantly, this DNA accessibility is determined by the 3-dimensional structure and how the DNA is packed in the nucleus[72]. Alterations in the DNA structure, but not sequences are known as epigenetic changes and describes that these structural changes can be heritable to daughter cells and offspring[73]. In

humans, DNA is packed into chromatin in which the DNA is wrapped around histones. DNA with an open chromatin structure is called euchromatin and is associated with active gene transcription[74]. DNA with a dense chromatin structure is known as heterochromatin and is associated with gene silencing or gene repression[74]. An important component of the epigenetic landscape is formed by the histones which are modulators of the chromatin structure. Histones consist of 5 families, the linker histone H1 and 4 core histones: H2A, H2B, H3 and H4. Eight histones (2 of each 4 core histones) form together a nucleosome wrapping approximately 150bp of DNA. The N-terminal histone tails stick out making them accessible for modifications. These modification can steer both transcriptional repression and activation dependent on the type of modification, see table I[75–77].

Table I. Histone modifications and their role on gene transcription

Functional association	Modification	Modification site
Gene activation	Acetylation	H3K9, H3K14, H3K18, H3K27, H3K56, H4K5, H4K8, H4K12, H4K16, H2BK6, H2BK7, H2BK16, H2BK17.
	Methylation	H3K4me1, H3K4me2, H3K4me3, H3K36me3, H3K79me2.
	Phosphorylation	H3S10ph.
Gene repression	Methylation	H3K9me2, H3K9me3, H3K27me2, H3K27me3, H4K20me3.

K = Lysine
S = Serine
me1 = monomethylation
me2 = demethylation
me3 = trimethylation
ph = phosphorylation.

One of the best-studied modifications is the trimethyl-modification on the position 4 (lysine) of histone 3 (H3K4me3), which is associated with active promotors[77]. Many of these histone modifications regulate gene expression by interacting with other proteins. For example, H3K4me3 can interact with TFIID, a

protein involved in the initiation of gene transcription[78]. In contrast, histone modification H3K9me3 interacts with Heterochromatin Protein 1 (HP1) thereby silencing gene transcription[79]. Histone modifications are generated by histone methylation transferases (HMTs) and histone acetylation transferases (HATs) and can be removed by histone deacetylases (HDACs) and histone demethylases (KDMs)[80,81].

Several studies have studied the role of epigenetic changes in autoimmune diseases including RA and SSc[82]. In RA, especially synovial fibroblasts (SFs) seem to display epigenetic alterations in patients[83–85]. In RA-specific SFs global hypomethylation correlates with increased levels of multiple receptors, adhesion molecules, and matrix-degrading enzymes and correlated with an activated phenotype[84]. Moreover, changes in DNA methylation have been observed in various immune cells of RA patients. The promotor of *CD40LG* in CD4+ T-cells and the promotor of *IL6* in B-cells are hypomethylated in RA patients[86,87].

Furthermore, behaviour and levels of the enzymes modifying the histones have been investigated in RA patients. Gillespie *et al*. showed that PBMCs isolated from RA patients exhibit enhanced histone deacetylase (HDAC) activity and inhibition of this class of enzymes showed the potential to reduce IL6 and TNFα proteins in a cell type and compound-dependent manner[88]. Moreover, levels of EZH2 (Enhancer of Zeste Homolog 2), a histone methyl transferase creating H3K27Me3 are increased in synovial fibroblasts isolated from RA patients[89]. Finally, in mouse models beneficial effects such as, reduced joint swelling, inflammation and cartilage destruction, have been obtained using inhibitors of HDACs[90,91].

In SSc patients, altered DNA methylation influenced the expression of both immune and fibrotic genes including *CD40L*, *TNFSF7*, *CD11a* and *FLI-1*[92–96]. Moreover, a genome-wide study identified several collagen genes both methylation and differential expressed in dermal fibroblasts[97].

Finally, histone modifying enzymes have been investigated as potential therapeutic targets in SSc. Inhibition of HDAC7 showed reduced cytokine-induced production of type I and type III collagen[98], whereas interfering with other histone modifying enzymes was found to reduce the accumulation of extracellular matrix in SSc mouse models[99].

Histone modifications play an important role in the regulation of immune-responsive genes as many immune-related genes are under control of epigenetic mechanisms. Various studies have been performed investigating which epigenetic marks play a role in the regulation of immune cells[100,101]. These changes seem specific for pathogens and environmental stimuli. For example stimulation of monocytes by either LPS (bacterial origin) or B-glucan (yeast origin) induces a different response leaving different epigenetic traces[102,103]. Interestingly, B-glucan exposure of monocytes induces long-lasting epigenetic changes[104]. Upon restimulation of these cells, induced cytokine production is observed. This phenomenon is called 'trained immunity' and the identified epigenetic changes include increased H3K4me3 levels and reduced H3K27me3 levels on genes that are enriched for immunological pathways[104]. These findings indicate that these genes are more easily accessible and ready for transcription upon new activation.

Besides regulation at the transcriptional level, gene expression is also regulated at the RNA level. Specifically, a subset of small non-coding RNAs have been discovered as regulator of RNA transcripts known as miRNAs[105,106]. miRNAs are small RNA molecules that can bind mRNA (often in the 3' UTR regions of mRNAs) thereby blocking translation and accelerating degradation of the mRNA in a process called RNA interference[107]. In short, miRNA are processed transcripts of ~20 nucleotides that bind complementary RNA often including several mismatches. Upon binding, the RISC complex can recognize the double stranded RNA and the targeted mRNA gets degraded by Dicer, an enzyme capable of cleaving double stranded RNA[107]. The opportunity that mRNA levels can be regulated qualifies miRNA as possible therapeutic targets in autoimmunity. For example, miRNA29 a key regulator of collagen expression was found decreased in patients with systemic sclerosis[108]. More examples of deregulated miRNAs are reviewed in[82,109]. Besides these small non-coding RNAs, other non-coding RNAs have surfaced as new players in disease and development.

Long non-coding RNAs and their relevance to rheumatic diseases

The human genome encompasses roughly 60,000 genes. As approximately 20,000 genes encode proteins, a larger proportion is of non-coding nature. Non-coding genes can be divided in groups of small and long non-coding RNAs, figure

3A. Besides a role of small non-coding RNAs (eg. miRNAs) in autoimmune disease, long non-coding RNAs have been linked to functions in immunity and autoimmune diseases[110]. lncRNAs are RNA transcripts with low or no coding potential with a length of over 200 nucleotides[111]. lncRNAs are often polyadenylated, and lack open reading frames (ORFs)[111,112].

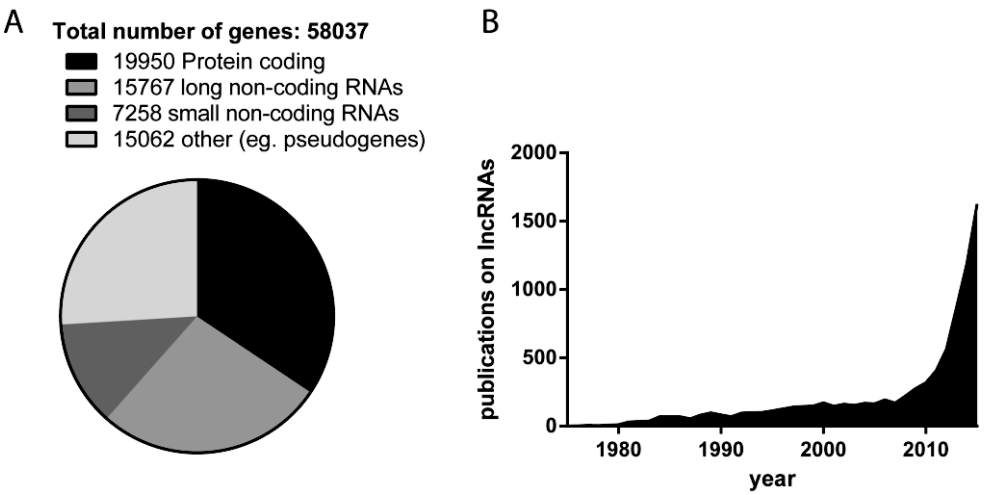

Figure 3. (**A**) The total number of genes divided into protein-coding genes, long non-coding RNAs (length of over 200 nucleotides), small non-coding RNAs and others (for example pseudogenes and immunoglobulin/T-cell receptor segments). Data was obtained from Gencode Version 25 (March 2016 freeze, GRCh38) - Ensembl 87. (**B**). Number of publications per year using search-term "Long non-coding RNAs", data subtracted from PubMed.

Since 2010, research and the number of publications studying the role of lncRNAs have exponentially increased (figure 3B). From these studies it has emerged that lncRNAs are important regulators of tissue physiology and disease processes[113]. Overall, lncRNAs are expressed at lower levels compared to coding genes and one of the most striking differences with coding RNAs is a more tissue-specific expression of lncRNAs[114,115]. Because of this tissue-specific expression, lncRNAs are considered important regulators of tissue specific physiology during development and during life[113,116,117]. It is key that during development gene expression is tightly regulated to prevent malformation of tissues and organs. For example, Sox2 is a transcription factor important for pluripotency and neuronal

development. Deregulation of Sox2 levels leads to malformation of neural tissues such as aberrant eye and brain development[114,118–121]. A lncRNA known as *Sox2* overlapping transcript (*Sox2ot*) is thought to play a role by safeguarding levels of Sox2 during neural development[122–124]. Humans have the highest count of lncRNAs and the number of lncRNAs has been shown to correlate with organism complexity[125,126]. Besides mammalian cells, lncRNAs are found in plants, yeast and even bacteria[127–129]. Although lncRNAs are found in many organisms, they are poorly conserved across species[115,130]. For example, only 14% of the mouse lncRNAs have a human orthologue[131], while another study shows that 12% of human lincRNAs have orthologous transcripts in other species[115]. Various types of lncRNAs have been described and can be divided into long intergenic non-coding RNAs (lincRNAs), intronic long non-coding RNAs, sense lncRNAs and antisense lncRNAs, figure 4.

Figure 4. Types of lncRNAs divided into intronic lncRNAs, intergenic lncRNAs, sense lncRNAs and antisense lncRNAs.

In comparison with small non-coding RNAs, the longer length of lncRNAs allows additional folding properties as stability and interaction abilities with DNA, RNA and protein[132]. lncRNAs can interact with single stranded DNA by direct base pairing or with double stranded DNA via triplex RNA-DNA structures[133]. lncRNAs are typically coexpressed with their neighbouring genes and are thought be involved with various gene regulatory processes[115]. Examples of lncRNAs that interfere with transcriptional and translational processes are summarized in figure 5[134,135]. Transcriptionally, lncRNAs are potent guiding molecules because of their ability to bind both DNA and proteins. By doing so, lncRNAs can bind co-activating or repressing proteins to specific genes and loci. Moreover, lncRNA have shown to play an important role in the modulation of epigenetic marks by recruiting histone modifying enzymes[136,137].

The other way around, lncRNAs can decoy proteins thereby preventing the interaction with specific regions[138–141]. Finally, expression of lncRNAs alone can result in interference with transcription of other genes. Transcriptional overlap of lncRNA *Airn*, but not its RNA transcripts were found crucial for silencing *Igf2r*[142]. Similarly, some lncRNAs are thought to function by disrupting DNA loops[143], while other lncRNAs (like Dum and HOTTIP) have been shown to establish DNA loops and coordinate gene expression[144,145].

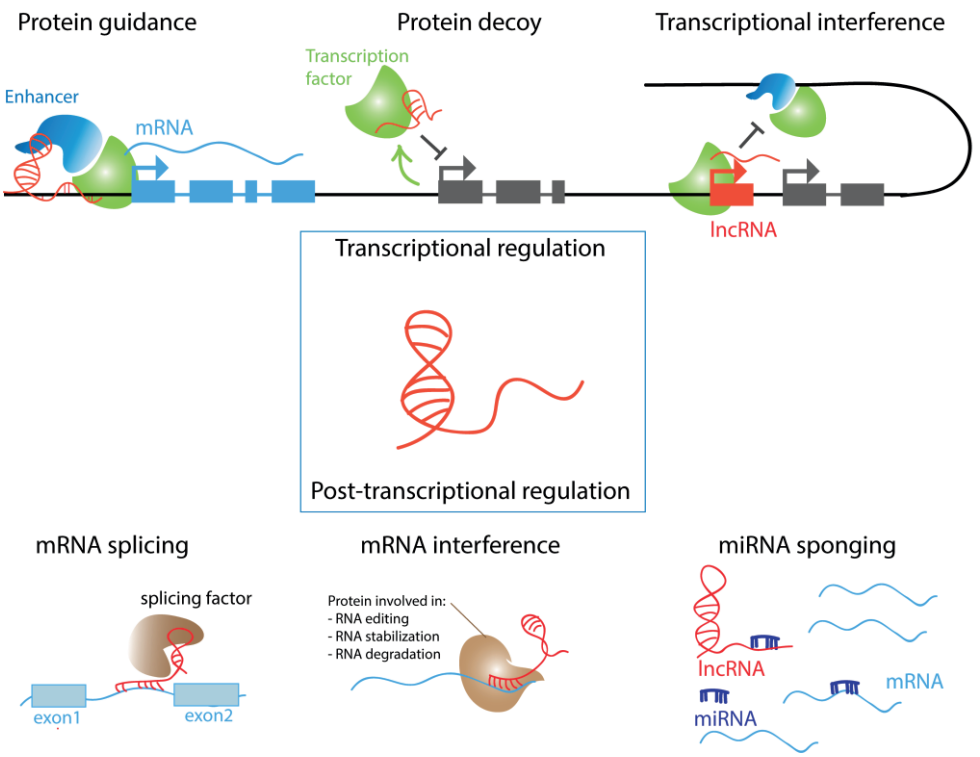

Figure 5. Mechanisms and function by lncRNAs. lncRNAs have potential roles in the regulation of transcription by protein guidance, protein decoy or via transcriptional interference (also mediated via DNA looping). Regulation of post-transcriptional processes by lncRNAs include the interference with mRNA splicing, stability, degradation, RNA editing and its interaction with miRNAs.

In addition, lncRNAs can regulate processes at the post-transcriptional level including, mRNA splicing, and interaction with mRNAs and miRNAs[146–148]. For example, ZEB2-AS1 and BACE1-AS1 are two antisense genes that interact with

mRNA by either influencing splicing or mRNA stability[149,150]. Antisense RNA genes are one of the largest subtypes of lncRNAs. These antisense RNA transcripts are transcribed from the opposite DNA strand of a sense gene in either close proximity or by partly overlapping, figure 4. The most prominent form of antisense transcription in the mammalian genome is of non-protein-coding nature[151]. Particularly, antisense ncRNAs represent an interesting subset as they are often involved in the regulation of its sense counterpart forming sense-antisense (SAS) gene pairs[152]. Besides ZEB2-AS1 and BACE1-AS, other SAS gene pairs have been shown to have implications in disease and development[153–156].

Studies with immune cells have shown that lncRNAs are important during the differentiation of immune cells[157,158]. Moreover, lncRNAs also play a role in differentiated immune cells and influence processes involved in innate and adaptive immunity[159–163]. Particularly in innate immunity lncRNA have been identified as important regulators. Many lncRNAs are under immunoregulatory control as shown by their responsiveness to external immune stimuli[164,165]. In both mouse and human derived monocytes/macrophages lncRNAs are responsive to LPS stimulation[156,165–167]. One of these studies shows that many of the lncRNA are coregulated or coexpressed with neighbouring protein-coding genes including *Nfkb2* and *Rel* two genes involved in NFKB-signalling[166]. Moreover, in human monocytes, 182 lncRNAs were induced by LPS, amongst others two lncRNAs in the *IL1B* locus that were identified as regulators of *IL1B* transcription and protein release[167]. Similarly, in monocyte-like cell lines, lncRNAs expression can be induced by various immune-stimuli[168,169]. For example, stimulation of THP1 cells with LPS showed 1161 lncRNA with differential expression[168]. Further knockdown experiments for some of these immune responsive-lncRNAs revealed their involvement with TNFα and IL6 levels[169]. Together, these studies show that lncRNAs have functional roles in our immune system and that they can influence the release of a variety of cytokines.

Besides these functional studies, genetic studies have linked lncRNAs to immunity and autoimmune diseases. Multiple GWAS have been performed to identify functional genetic regions that contribute to autoimmunity. Many of these associated regions are located in loci without coding genes, however can contain unannotated non-coding genes[164,170,171]. Nonetheless, SNPs occurring in these non-coding genes could have functional effects and subsequently result in

phenotypic changes contributing to complex diseases, including cancers and AID[44,172,173]. SNPs can strongly affect gene expression and are also known as expression quantitative trait loci (*cis*-eQTL)[174,175]. Many of these variants interfere with the binding of transcription factors or enhancers, however also lncRNAs could be functionally affected[176,177]. Several studies have linked genetic variants to changes in lncRNA levels through eQTL studies. Almhof *et al*. investigated cis-acting genetic variation that regulate expression of lncRNAs in monocytes isolated from 188 healthy donors[178]. 258 lncRNAs were detected with at least one associated cis-regulatory SNP of which 20% co-regulated with the closest protein coding gene[178]. Not only expression levels could be affected by these SNPs but studies have shown that genetic variants influence the predicted structure of lncRNAs related to immune diseases as T1D and inflammatory bowel disease[179]. Together, these studies show that associated variants can affect non-coding genes by altering the RNA stability, RNA expression levels and RNA structure and are therefore important candidates for further investigation in disease pathogenesis.

Outline of this thesis

Multiple factors have been identified that contribute to autoimmunity. Studies like association studies, gene expression studies and familial studies have revealed that both environmental and genetic factors contribute to disease pathogenesis. Although a dominant role for HLA genes in autoimmunity is very likely, non-HLA protein encoding genes have been identified as important clinical drug targets as well. In RA, anti-TNFα and anti-IL6 therapies are two successful strategies to treat patients. However, the mechanism by which TNFα and IL6 are upregulated in RA patients is yet unknown. In **Chapter 2** of this thesis, transcriptional regulation and the protein levels of IL6 and TNFα were investigated in RA patients. In particular, this study addressed whether epigenetic changes mediate enhanced IL6 and TNFα levels in early untreated RA patients and if these epigenetic changes can be detected in circulating monocytes isolated from these patients.

More recently, it was postulated that besides coding genes, non-coding genes also possess functional roles in development, immunity and various diseases. However, the involvement of these non-coding genes in autoimmunity and their

mode of action is poorly understood. **Chapter 3** of this thesis summarizes how both coding and non-coding variants can affect genes and how these variants may contribute to RA and other autoimmune diseases[44]. Interestingly, the majority of risk variants locate to non-coding regions of which the underlying disease-contributing mechanism is often unknown[180]. A RA-associated locus with many non-coding variants is the *TRAF1-C5* risk locus located on chromosome 9[181,182]. As described in **Chapter 4 and 5**, none of the identified variants are located in known coding regions but instead are synonymous, intronic or located in intergenic and UTR regions[183]. Such variants might also cover non-coding genes which are now hypothesized as a novel group of candidate genes that might explain part of the genetic variants associated with disease pathogenesis[164,172]. In addition, **Chapter 5** describes the identification of such a novel disease candidate gene, of non-coding nature, in the *TRAF1-C5* locus[164].

Besides a role for lncRNAs in RA, also lncRNAs are proposed to play pathogenic roles in SSc[184]. Investigating deregulated non-coding genes in SSc patients may lead to the identification of novel SSc biomarker genes and potentially new druggable targets. In **Chapter 6** of this thesis, expression of both coding and non-coding genes of SSc patients was compared to healthy controls to identify such targets. Moreover, the molecular mechanism by which deregulated antisense RNA genes play a role in SSc was investigated particularly. In **Chapter 7** of this thesis, the molecular mechanism of another lncRNA known as Sox2ot was further investigated. Sox2ot is a non-coding RNA that is located near Sox2, an important transcription factor for pluripotency and neural development. Together, these studies aid in unravelling the function of lncRNAs and to understand their role and involvement in development and disease pathogenesis. Finally, general remarks, implication and future directions are discussed in **Chapter 8**. To summarize, in this thesis we aimed at identifying and unravelling of enigmatic transcriptional mechanisms that contribute to rheumatoid arthritis (RA) but also to systemic sclerosis (SSc) with a particular focus on non-coding genes.

Answering the following questions was of main interest in this thesis:

i.) Are epigenetic changes underlying the development of RA and can they be detected in early untreated RA patients?

ii.) What are the mechanisms by which genetic variation contributes to autoimmunity and how do non-coding genes play a role in this process?

iii.) Are non-coding genes deregulated in AID and how can these deregulated genes contribute to autoimmunity?

iv.) What are the molecular mechanisms by which non-coding genes can function and play a role in disease and development?

References

1 R0ichard-Miceli C, Criswell LA. Emerging patterns of genetic overlap across autoimmune disorders. *Genome Med* 2012; **4**: 1–9.

2 Somers EC, Thomas SL, Smeeth L, Hall AJ. Autoimmune diseases co-occurring within individuals and within families: a systematic review. *Epidemiology* 2006; **17**: 202–17.

3 Cooper GS, Stroehla BC. The epidemiology of autoimmune diseases. Autoimmun. Rev. 2003; **2**: 119–125.

4 Barnes J, Mayes MD. Epidemiology of systemic sclerosis: incidence, prevalence, survival, risk factors, malignancy, and environmental triggers. *Curr Opin Rheumatol* 2012; **24**: 165–70.

5 van Vollenhoven RF. Sex differences in rheumatoid arthritis: more than meets the eye... *BMC Med* 2009; **7**: 12.

6 Kivity S, Ehrenfeld M. Can we explain the higher prevalence of autoimmune disease in women? *Expert Rev Clin Immunol* 2010; **6**: 691–4.

7 Amur S, Parekh A, Mummaneni P. Sex differences and genomics in autoimmune diseases. *J Autoimmun* 2012; **38**: J254-65.

8 Tiniakou E, Costenbader KH, Kriegel M a. Sex-specific environmental influences on the development of autoimmune diseases. *Clin Immunol* 2013; **149**: 182–91.

9 Zhernakova A, Withoff S, Wijmenga C. Clinical implications of shared genetics and pathogenesis in autoimmune diseases. *Nat Rev Endocrinol* 2013; **9**: 646–59.

10 Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C *et al.* Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet* 2011; **7**: e1002254.

11 Farh KK-H, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 2015; **518**: 337–43.

12 Danzer C, Mattner J. Impact of microbes on autoimmune diseases. *Arch Immunol Ther Exp (Warsz)* 2013; **61**: 175–86.

13 Barragán-Martínez C, Speck-Hernández C a, Montoya-Ortiz G, Mantilla RD, Anaya J-M, Rojas-Villarraga A. Organic solvents as risk factor for autoimmune diseases: a systematic review and meta-analysis. *PLoS One* 2012; **7**: e51506.

14 Draborg AH, Duus K, Houen G. Epstein-Barr virus in systemic autoimmune diseases. *Clin Dev Immunol* 2013; **2013**: 535738.

15 Wang L, Wang F-S, Gershwin ME. Human autoimmune diseases: a comprehensive update. *J Intern Med* 2015; **278**: 369–95.

16 Alamanos Y, Voulgari P V, Drosos A a. Incidence and prevalence of rheumatoid arthritis, based on the 1987 American College of Rheumatology criteria: a systematic review. *Semin Arthritis Rheum* 2006; **36**: 182–8.

17 Smolen JS, Steiner G. Therapeutic strategies for rheumatoid arthritis. *NatRev Drug Discov* 2003; **2**: 473–488.

18 Smolen JS, Aletaha D, McInnes IB. Rheumatoid arthritis. Lancet. 2016; **388**: 2023–2038.

19 Sokka T, Toloza S, Cutolo M, Kautiainen H, Makinen H, Gogus F *et al.* Women, men, and rheumatoid arthritis: analyses of disease activity, disease characteristics, and treatments in the QUEST-RA study. *Arthritis Res Ther* 2009; **11**: R7.

20 Aletaha D, Alasti F, Smolen JS. Rheumatoid factor, not antibodies against citrullinated proteins, is associated with baseline disease activity in rheumatoid arthritis clinical trials. *Arthritis Res Ther* 2015; **17**: 229.

21 van Gaalen F a, van Aken J, Huizinga TWJ, Schreuder GMT, Breedveld FC, Zanelli E *et al.* Association between HLA class II genes and autoantibodies to cyclic citrullinated peptides (CCPs) influences the severity of rheumatoid arthritis. *Arthritis Rheum* 2004; **50**: 2113–21.

22 Brink M, Verheul MK, Rönnelid J, Berglin E, Holmdahl R, Toes REM *et al.* Anti-carbamylated protein antibodies in the pre-symptomatic phase of rheumatoid arthritis, their relationship with multiple anti-citrulline peptide antibodies and association with radiological damage. *Arthritis Res Ther* 2015; **17**: 25.

23 Shi J, Knevel R, Suwannalai P, van der Linden MP, Janssen GMC, van Veelen P a *et al.* Autoantibodies recognizing carbamylated proteins are present in sera of patients with rheumatoid arthritis and predict joint damage. *Proc Natl Acad Sci U S A* 2011; **108**: 17372–7.

24 Gonzalez A, Icen M, Kremers HM, Crowson CS, Davis JM, Therneau TM *et al.* Mortality trends in rheumatoid arthritis: The role of rheumatoid factor. *J Rheumatol* 2008; **35**: 1009–1014.

25 Nell VPK, Machold KP, Stamm T a, Eberl G, Heinzl H, Uffmann M *et al.* Autoantibody profiling as early diagnostic and prognostic tool for rheumatoid arthritis. *Ann Rheum Dis* 2005; **64**: 1731–6.

26 Suzuki K, Sawada T, Murakami a, Matsui T, Tohma S, Nakazono K *et al.* High diagnostic performance of ELISA detection of antibodies to citrullinated antigens in rheumatoid arthritis. *Scand J Rheumatol* 2009; **32**: 197–204.

27 Jiang X, Trouw L a, van Wesemael TJ, Shi J, Bengtsson C, Källberg H *et al.* Anti-CarP antibodies in two large cohorts of patients with rheumatoid arthritis and their relationship to genetic risk factors, cigarette smoking and other autoantibodies. *Ann Rheum Dis* 2014; **73**: 1761–8.

28 Krol a, Garred P, Heegaard NHH, Christensen a F, Hetland ML, Stengaard-Pedersen K *et al.* Interactions between smoking, increased serum levels of anti-CCP antibodies, rheumatoid factors, and erosive joint disease in patients with early, untreated rheumatoid arthritis. *Scand J Rheumatol* 2015; **44**: 8–12.

29 Shi J, van de Stadt L a, Levarht EWN, Huizinga TWJ, Hamann D, van Schaardenburg D *et al.* Anti-carbamylated protein (anti-CarP) antibodies precede the onset of rheumatoid arthritis. *Ann Rheum Dis* 2014; **73**: 780–3.

30 Shi J, van Steenbergen HW, van Nies J a. B, Levarht EWN, Huizinga TWJ, van der Helm-van Mil AHM *et al.* The specificity of anti-carbamylated protein antibodies for rheumatoid arthritis in a setting of early arthritis. *Arthritis Res Ther* 2015; **17**: 339.

31 Wei S-T, Sun Y-H, Zong S-H, Xiang Y-B. Serum Levels of IL-6 and TNF-α May Correlate with Activity and Severity of Rheumatoid Arthritis. *Med Sci Monit* 2015; **21**: 4030–4038.

32 Houssiau FA, Devogelaer J, Damme JOVAN, Deuxchaisnes CNDE, Snick JVAN. Interleukin-6 in synovial fluid and serum of patients with rheumatoid arthritis and other inflammatory arthritides. *Arthritis Rheum* 1988; **31**: 784–788.

33 Tetta C, Camussi G, Modena V, Vittorio C Di, Baglioni C. Tumour necrosis factor in serum and synovial fluid of patients with active and severe rheumatoid arthritis. *Ann Rheum Dis* 1990; **49**: 665–667.

34 Xu S. TNF inhibitor therapy for rheumatoid arthritis. *Biomed Reports* 2012; **1**: 177–184.

35 Woodrick R, Ruderman EM. Anti-Interleukin-6 Therapy in Rheumatoid Arthritis. 2010; **68**.

36 Ma X, Xu S. TNF inhibitor therapy for rheumatoid arthritis. *Biomed reports* 2013; **1**: 177–184.

37 Srirangan S, Choy EH. The role of interleukin 6 in the pathophysiology of rheumatoid arthritis. *Ther Adv Musculoskelet Dis* 2010; **2**: 247–56.

38 Hennigan S, Kavanaugh A. Interleukin-6 inhibitors in the treatment of rheumatoid arthritis. *Ther Clin Risk Manag* 2008; **4**: 767–775.

39 Macgregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K *et al.* Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 2000; **43**: 30–37.

40 Frisell T, Holmqvist M, Källberg H, Klareskog L, Alfredsson L, Askling J. Familial risks and heritability of rheumatoid arthritis: role of rheumatoid factor/anti-citrullinated protein antibody status, number and type of affected relatives, sex, and age. *Arthritis Rheum* 2013; **65**: 2773–82.

41 Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2013. doi:10.1038/nature12873.

42 Gregersen PK, Silver J, Winchester RJ. The Shared Epitope Hypothesis. *Arthritis Rheum* 1987; **30**: 1205–1212.

43 Scally SW, Petersen J, Law SC, Dudek NL, Nel HJ, Loh KL *et al.* A molecular basis for the association of the HLA-DRB1 locus, citrullination, and rheumatoid arthritis. *J Exp Med* 2013; **210**: 2569–82.

44 Messemaker TC, Huizinga TW, Kurreeman F. Immunogenetics of rheumatoid arthritis: Understanding functional implications. *J Autoimmun* 2015; **64**: 7–14.

45 Rawlings DJ, Dai X, Buckner JH, Alerts E. The Role of PTPN22 Risk Variant in the Development of Autoimmunity: Finding Common Ground between Mouse and Human. *J Immunol* 2015; **194**: 2977–84.

46 Zhang J, Zahir N, Jiang Q, Miliotis H, Heyraud S, Meng X *et al.* The autoimmune disease – associated PTPN22 variant promotes calpain-mediated Lyp / Pep degradation associated with lymphocyte and dendritic cell hyperresponsiveness. *Nat Genet* 2011; **43**: 902–907.

47 Rieck M, Arechiga A, Onengut-gumuscu S, Greenbaum C, Concannon P, Buckner JH *et al.* Genetic variation in PTPN22 corresponds to altered function of T and B lymphocytes. *J Immunol* 2007; **179**: 4704–4710.

48 Al MASET. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. *Arthritis Rheum* 1980; **23**: 581–90.

49 Steen D V. Epidemiology of systemic sclerosis. *Rheumatology* 2003; **12**: 1455–1462.

50 Allanore Y, Simms R, Distler O, Trojanowska M, Pope J, Denton CP *et al.* Systemic sclerosis. *Nat Rev Dis Prim* 2015; **1**: 15002.

51 Sunderkötter C, Riemekasten G. Pathophysiology and clinical consequences of Raynaud's phenomenon related to systemic sclerosis. *Rheumatology (Oxford)* 2006; **45 Suppl 3**: iii33-5.

52 Hachulla E, Launay D. Diagnosis and classification of systemic sclerosis. *Clin Rev Allergy Immunol* 2011; **40**: 78–83.

53 Steen VD. Autoantibodies in systemic sclerosis. *Semin Arthritis Rheum* 2005; **35**: 35–42.

54 Lomeo RM. Progressive Systemic Sclerosis Sine Scleroderma Presenting as Pulmonary Interstitial Fibrosis. *Am J Med* 1989; **87**: 525–527.

55 De Martinis M, Ciccarelli F, Sirufo MM, Ginaldi L. An overview of environmental risk factors in systemic sclerosis. *Expert Rev Clin Immunol* 2016; **12**: 465–78.

56 Dospinescu P, Jones GT, Basu N. Environmental risk factors in systemic sclerosis. *Curr Opin Rheumatol* 2013; **25**: 179–83.

57 Korman BD, Criswell L a. Recent advances in the genetics of systemic sclerosis: toward biological and clinical significance. *Curr Rheumatol Rep* 2015; **17**: 21.

58 Mayes MD. The genetics of scleroderma: looking into the postgenomic era. *Curr Opin Rheumatol* 2012; **24**: 677–84.

59 Zochling J, Newell F, Charlesworth JC, Leo P, Stankovich J, Cortes A *et al.* An Immunochip-based interrogation of scleroderma susceptibility variants identifies a novel association at DNASE1L3. *Arthritis Res Ther* 2014; **16**: 438.

60 Gilchrist FC, Bunn C, Foley PJ, Lympany PA, Black CM, Welsh KI *et al.* Class II HLA associations with autoantibodies in scleroderma: a highly significant role for HLA-DP. *Genes Immun* 2001; **2**: 76–81.

61 Whitfield ML, Finlay DR, Murray JI, Troyanskaya OG, Chi J, Pergamenschikov A *et al.* Systemic and cell type-specific gene expression patterns in scleroderma skin. *Proc Natl Acad Sci U S A* 2003; **100**: 12319–12324.

62 Assassi S, Swindell WR, Wu M, Tan FD, Khanna D, Furst DE *et al.* Dissecting the heterogeneity of skin gene expression patterns in systemic sclerosis. *Arthritis Rheumatol (Hoboken, NJ)* 2015; **67**: 3016–26.

63 Gardner H, Shearstone JR, Bandaru R, Crowell T, Lynes M, Trojanowska M *et al.* Gene profiling of scleroderma skin reveals robust signatures of disease that are imperfectly reflected in the transcript profiles of explanted fibroblasts. *Arthritis Rheum* 2006; **54**: 1961–73.

64 Milano A, Pendergrass S a, Sargent JL, George LK, McCalmont TH, Connolly MK *et al.* Molecular subsets in the gene expression signatures of scleroderma skin. *PLoS One* 2008; **3**: e2696.

65 Pendergrass S a, Hayes E, Farina G, Lemaire R, Farber HW, Whitfield ML *et al.* Limited systemic sclerosis patients with pulmonary arterial hypertension show biomarkers of inflammation and vascular injury. *PLoS One* 2010; **5**: e12106.

66 Christmann RB, Hayes E, Pendergrass S, Padilla C, Farina G, Affandi AJ *et al.* Interferon and alternative activation of monocyte/macrophages in systemic sclerosis-associated pulmonary arterial hypertension. *Arthritis Rheum* 2011; **63**: 1718–28.

67 Johnson ME, Pioli P a, Whitfield ML. Gene expression profiling offers insights into the role of innate immune signaling in SSc. *Semin Immunopathol* 2015; **37**: 501–9.

68 Bournia VKK, Vlachoyiannopoulos PG, Selmi C, Moutsopoulos HM, Gershwin ME. Recent advances in the treatment of systemic sclerosis. *Clin Rev Allergy Immunol* 2009; **36**: 176–200.

69 Taroni JN, Martyanov V, Mahoney JM, Whitfield ML. A functional genomic meta-analysis of clinical trials in systemic sclerosis: towards precision medicine and combination therapy. *J Invest Dermatol* 2016. doi:10.1016/j.jid.2016.12.007.

70 Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F *et al.* GENCODE: The reference human genome annotation for the ENCODE project. *Genome Res* 2012; **22**: 1760–1774.

71 Jacob F, Monod J. Genetic Regulatory Mechanisms in the Synthesis. *J Mol Miol* 1960; **3**: 318–356.

72 Song L, Crawford GE. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc* 2010; **2010**: 1–13.

73 Morgan HD, Sutherland HG, Martin DI, Whitelaw E. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* 1999; **23**: 314–318.

74 Hood L, Galas D. The digital code of DNA. *Nature* 2003; **421**: 444–8.

75 Tessarz P, Kouzarides T. Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol* 2014; **15**: 703–8.

76 Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res* 2011; **21**: 381–95.

77 Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* 2007; **129**: 823–37.

78 Lauberth SM, Nakayama T, Wu X, Ferris AL, Tang Z, Hughes SH *et al.* H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* 2013; **152**: 1021–36.

79 Boros J, Arnoult N, Stroobant V, Collet J-F, Decottignies A. Polycomb repressive complex 2 and H3K27me3 cooperate with H3K9 methylation to maintain heterochromatin protein 1α at chromatin. *Mol Cell Biol* 2014; **34**: 3662–74.

80 Wendler DS. Problems with the consensus definition of the therapeutic misconception. *J Clin Ethics* 2013; **24**: 387–394.

81 Kelley RK, Wang G, Venook AP. Biomarker use in colorectal cancer therapy. JNCCN J. Natl. Compr. Cancer Netw. 2011; **9**: 1293–1303.

82 Aslani S, Mahmoudi M, Karami J, Jamshidi AR, Malekshahi Z, Nicknam MH. Epigenetic alterations underlying autoimmune diseases. *Autoimmunity* 2016; **49**: 69–83.

83 Takami N, Osawa K, Miura Y, Komai K, Taniguchi M, Shiraishi M *et al.* Hypermethylated promoter region of DR3, the death receptor 3 gene, in rheumatoid arthritis synovial cells. *Arthritis Rheum* 2006; **54**: 779–87.

84 Karouzakis E, Gay RE, Michel B a, Gay S, Neidhart M. DNA hypomethylation in rheumatoid arthritis synovial fibroblasts. *Arthritis Rheum* 2009; **60**: 3613–22.

85 Neidhart M, Rethage J, Kuchen S, Ku P, Billingham ME, Gay RE *et al.* Retrotransposable l1 elements expressed in rheumatoid arthritis synovial tissue. *Arthritis Rheum* 2000; **43**: 2634–2647.

86 Nile CJ, Read RC, Akil M, Duff GW, Wilson AG. Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis. *Arthritis Rheum* 2008; **58**: 2686–93.

87 Liao J, Liang G, Xie S, Zhao H, Zuo X, Li F *et al.* CD40L demethylation in CD4(+) T cells from women with rheumatoid arthritis. *Clin Immunol* 2012; **145**: 13–8.

88 Gillespie J, Savic S, Wong C, Hempshall A, Inman M, Emery P *et al.* Histone deacetylases are dysregulated in rheumatoid arthritis and a novel histone deacetylase 3-selective inhibitor reduces interleukin-6 production by peripheral blood mononuclear cells from rheumatoid arthritis patients. *Arthritis Rheum* 2012; **64**: 418–22.

89 Trenkmann M, Brock M, Gay RE, Kolling C, Speich R, Michel B a *et al.* Expression and function of EZH2 in synovial fibroblasts: epigenetic repression of the Wnt inhibitor SFRP1 in rheumatoid arthritis. *Ann Rheum Dis* 2011; **70**: 1482–8.

90 Nishida K, Komiyama T, Miyazawa S-I, Shen Z-N, Furumatsu T, Doi H *et al.* Histone deacetylase inhibitor suppression of autoantibody-mediated arthritis in mice via regulation of p16INK4a and p21(WAF1/Cip1) expression. *Arthritis Rheum* 2004; **50**: 3365–76.

91 Manabe H, Nasu Y, Komiyama T, Furumatsu T, Kitamura A, Miyazawa S *et al.* Inhibition of histone deacetylase down-regulates the expression of hypoxia-induced vascular endothelial growth factor by rheumatoid synovial fi broblasts. *Inflamm Res* 2008; **57**: 4–10.

92 Czuwara-Ladykowska J, Shirasaki F, Jackers P, Watson DK, Trojanowska M. Fli-1 inhibits collagen type I production in dermal fibroblasts via an Sp1-dependent pathway. *J Biol Chem* 2001; **276**: 20839–48.

93 Jiang H, Xiao R, Lian X, Kanekura T, Luo Y, Yin Y *et al.* Demethylation of TNFSF7 contributes to CD70 overexpression in CD4+ T cells from patients with systemic sclerosis. *Clin Immunol* 2012; **143**: 39–44.

94 Lian X, Xiao R, Hu X, Kanekura T, Jiang H, Li Y *et al.* DNA demethylation of CD40l in CD4+ T cells from women with ystemic sclerosis: a possible explanation for female susceptibility. *Arthritis Rheum* 2012; **64**: 2338–45.

95 Wang Y, Shu Y, Xiao Y, Wang Q, Kanekura T, Li Y *et al.* Hypomethylation and overexpression of ITGAL (CD11a) in CD4(+) T cells in systemic sclerosis. *Clin Epigenetics* 2014; **6**: 25.

96 Wang Y, Fan P-S, Kahaleh B. Association between enhanced type I collagen expression and epigenetic repression of the FLI1 gene in scleroderma fibroblasts. *Arthritis Rheum* 2006; **54**: 2271–9.

97 Altorok N, Tsou P-S, Coit P, Khanna D, Sawalha AH. Genome-wide DNA methylation analysis in dermal fibroblasts from patients with diffuse and limited systemic sclerosis reveals common and subset-specific DNA methylation aberrancies. *Ann Rheum Dis* 2015; **74**: 1612–20.

98 Hemmatazad H, Rodrigues HM, Maurer B, Brentano F, Pileckyte M, Distler JHW *et al.* Histone deacetylase 7, a potential target for the antifibrotic treatment of systemic sclerosis. *Arthritis Rheum* 2009; **60**: 1519–29.

99 Huber LC, Distler JHW, Moritz F, Hemmatazad H, Hauser T, Michel B a *et al.* Trichostatin A prevents the accumulation of extracellular matrix in a mouse model of bleomycin-induced skin fibrosis. *Arthritis Rheum* 2007; **56**: 2755–64.

100 Reiner SL. Epigenetic control in the immune response. *Hum Mol Genet* 2005; **14 Spec No**: R41-6.

101 Lim PS, Li J, Holloway AF, Rao S. Epigenetic regulation of inducible gene expression in the immune system. *Immunology* 2013; **139**: 285–93.

102 Ifrim DC, Quintin J, Joosten L a B, Jacobs C, Jansen T, Jacobs L *et al.* Trained immunity or tolerance: opposing functional programs induced in human monocytes after engagement of various pattern recognition receptors. *Clin Vaccine Immunol* 2014; **21**: 534–45.

103 Saeed S, Quintin J, Kerstens HHD, Rao N a, Aghajanirefah A, Matarese F *et al.* Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science (80-)* 2014; **345**: 1251086.

104 Quintin J, Saeed S, Martens JH a, Giamarellos-Bourboulis EJ, Ifrim DC, Logie C *et al.* Candida albicans infection affords protection against reinfection via functional reprogramming of monocytes. *Cell Host Microbe* 2012; **12**: 223–32.

105 Fabian MR, Sonenberg N, Filipowicz W. Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* 2010; **79**: 351–79.

106 Afonso-Grunz F, Müller S. Principles of miRNA-mRNA interactions: beyond sequence complementarity. *Cell Mol Life Sci* 2015; **72**: 3127–41.

107 Jonas S, Izaurralde E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet* 2015; **16**: 421–33.

108 Maurer B, Stanczyk J, Jüngel A, Akhmetshina A, Trenkmann M, Brock M *et al.* MicroRNA-29, a key regulator of collagen expression in systemic sclerosis. *Arthritis Rheum* 2010; **62**: 1733–43.

109 Chen X-M, Huang Q-C, Yang S-L, Chu Y-L, Yan Y-H, Han L *et al.* Role of Micro RNAs in the Pathogenesis of Rheumatoid Arthritis: Novel Perspectives Based on Review of the Literature. *Medicine (Baltimore)* 2015; **94**: e1326.

110 Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet* 2011; **12**: 861–74.

111 Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 2012; **22**: 1775–89.

112 Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT *et al.* RNA Maps Reveal New RNA Classes and a Possible Function for Pervasive Transcription. *Science (80-)* 2007; **316**: 1484–1488.

113 Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. *Cell* 2013; **152**: 1298–307.

114 Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* 2015; **47**: 199–208.

115 Pj B, Cao H, Np C, Gk C, Davis S, Day N *et al.* Mohammad Heydarian thesis proposal - 1.19.2012 Page 27. *Genes …* 2012; **447**: 27–32.

116 Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet* 2014; **15**: 7–21.

117 van Leeuwen S, Mikkers H. Long non-coding RNAs: Guardians of development. *Differentiation* 2010; **80**: 175–83.

118 Taranova O V, Magness ST, Fagan BM, Wu Y, Surzenko N, Hutton SR *et al.* SOX2 is a dose-dependent regulator of retinal neural progenitor competence. *Genes Dev* 2006; **20**: 1187–1202.

119 Niu W, Zang T, Zou Y, Fang S, Smith DK, Bachoo R *et al.* In vivo reprogramming of astrocytes to neuroblasts in the adult brain. *Nat Cell Biol* 2013; **15**: 1164–75.

120 Bylund M, Andersson E, Novitch BG, Muhr J. Vertebrate neurogenesis is counteracted by Sox1-3 activity. *Nat Neurosci* 2003; **6**: 1162–8.

121 Graham V, Khudyakov J, Ellis P, Pevny L. SOX2 functions to maintain neural progenitor identity. *Neuron* 2003; **39**: 749–765.

122 Askarian-Amiri ME, Seyfoddin V, Smart CE, Wang J, Kim JE, Hansji H *et al.* Emerging role of long non-coding RNA SOX2OT in SOX2 regulation in breast cancer. *PLoS One* 2014; **9**: e102140.

123 Hou Z, Zhao W, Zhou J, Shen L, Zhan P, Xu C *et al.* A long noncoding RNA Sox2ot regulates lung cancer cell proliferation and is a prognostic indicator of poor survival. *Int J Biochem Cell Biol* 2014; **53**: 380–8.

124 Shahryari A, Rafiee MR, Fouani Y, Oliae NA, Samaei NM, Shafiee M *et al.* Two novel splice variants of SOX2OT, SOX2OT-S1, and SOX2OT-S2 are coupregulated with SOX2 and OCT4 in esophageal squamous cell carcinoma. *Stem Cells* 2014; **32**: 126–134.

125 Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 2014; **505**: 635–40.

126 Taft RJ, Pheasant M, Mattick JS. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* 2007; **29**: 288–99.

127 Chekanova J a. Long non-coding RNAs and their functions in plants. *Curr Opin Plant Biol* 2015; **27**: 207–16.

128 Yamashita A, Shichino Y, Yamamoto M. The long non-coding RNA world in yeasts. *Biochim Biophys Acta* 2016; **1859**: 147–54.

129 Weinberg Z, Perreault J, Meyer MM, Breaker RR. Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. *Nature* 2009; **462**: 656–9.

130 Chodroff RA, Goodstadt L, Sirey TM, Oliver PL, Davies KE, Green ED *et al.* Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biol* 2010; **11**: R72.

131 Church DM, Goodstadt L, Hillier LW, Zody MC, Goldstein S, She X *et al.* Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biol* 2009; **7**: e1000112.

132 Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. *Nature* 2012; **482**: 339–46.

133 Li Y, Syed J, Sugiyama H. RNA-DNA Triplex Formation by Long Noncoding RNAs. *Cell Chem Biol* 2016; **23**: 1325–1333.

134 Guil S, Esteller M. Cis-acting noncoding RNAs: friends and foes. *Nat Struct Mol Biol* 2012; **19**: 1068–75.

135 Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 2012; **81**: 145–66.

136 Richard I. Joh, Christina M. Palmieri, Ian T Hill and MM. Regulation of histone methylation by noncoding RNAs. *Biochim Biophys Acta* 2014; **1839**: 1385–1394.

137 Holdt LM, Hoffmann S, Sass K, Langenberger D, Scholz M, Krohn K *et al.* Alu elements in ANRIL non-coding RNA at chromosome 9p21 modulate atherogenic cell functions through trans-regulation of gene networks. *PLoS Genet* 2013; **9**: e1003588.

138 Xue Z, Hennelly S, Doyle B, Gulati A a, Novikova I V, Sanbonmatsu KY *et al.* A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage. *Mol Cell* 2016; **64**: 37–50.

139 Klattenhoff C a, Scheuermann JC, Surface LE, Bradley RK, Fields P a, Steinhauser ML *et al.* Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* 2013; **152**: 570–83.

140 Grote P, Herrmann BG. The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. *RNA Biol* 2013; **10**: 1579–1585.

141 Bläss G, Kellis M, Werber M, Herrmann BG. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell* 2013; **24**: 206–214.

142 Latos PA, Pauler FM, Koerner M V, Ba H, Hudson QJ, Stocsits RR *et al.* Airn Transcriptional Overlap, But Not Its lncRNA Products, Induces Imprinted Igf2r Silencing. *Science (80-)* 2012; **338**.

143 Quinodoz S, Guttman M. Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. *Trends Cell Biol* 2014; **24**: 651–63.

144 Wang L, Zhao Y, Bao X, Zhu X, Kwok YK-Y, Sun K *et al.* LncRNA Dum interacts with Dnmts to regulate Dppa2 expression during myogenic differentiation and muscle regeneration. *Cell Res* 2015; **25**: 335–50.

145 Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 2011; **472**: 120–4.

146 Thomson DW, Dinger ME. Endogenous microRNA sponges: evidence and controversy. *Nat Rev Genet* 2016; **17**: 272–83.

147 Bayoumi AS, Sayed A, Broskova Z, Teoh J-P, Wilson J, Su H *et al.* Crosstalk between Long Noncoding RNAs and MicroRNAs in Health and Disease. *Int J Mol Sci* 2016; **17**: 356.

148 Chen M-T, Lin H-S, Shen C, Ma Y-N, Wang F, Zhao H-L *et al.* PU.1-Regulated Long Noncoding RNA lnc-MC Controls Human Monocyte/Macrophage Differentiation through Interaction with MicroRNA 199a-5p. *Mol Cell Biol* 2015; **35**: 3212–24.

149 Lan T, Chang L, Wu L, Yuan Y. Downregulation of ZEB2-AS1 decreased tumor growth and metastasis in hepatocellular carcinoma. *Mol Med Rep* 2016; **14**: 4606–4612.

150 Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE *et al.* Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β-secretase. *Nat Med* 2008; **14**: 723–730.

151 Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M *et al.* Antisense transcription in the mammalian transcriptome. *Science (80-)* 2005; **309**: 1564–1566.

152 Chen J, Sun M, Hurst LD, Carmichael GG, Rowley JD. Genome-wide analysis of coordinate expression and evolution of human cis- encoded sense-antisense transcripts. *Trends Genet* 2005; **21**: 326–329.

153 Peng H, Liu Y, Tian J, Ma J, Tang X, Rui K *et al.* The Long Noncoding RNA IFNG-AS1 Promotes T Helper Type 1 Cells Response in Patients with Hashimoto's Thyroiditis. *Sci Rep* 2015; **5**: 17702.

154 Chen N, Guo D, Xu Q, Yang M, Wang D, Peng M. Long non-coding RNA FEZF1-AS1 facilitates cell proliferation and migration in colorectal carcinoma. *Oncotarget* 2016; **7**.

155 Liu C, Lin J. Long noncoding RNA ZEB1-AS1 acts as an oncogene in osteosarcoma by epigenetically activating ZEB1. *Am J Transl Res* 2016; **8**: 4095–4105.

156 Chan J, Atianand M, Jiang Z, Carpenter S, Aiello D, Elling R *et al.* Cutting Edge: A Natural Antisense Transcript, AS-IL1α, Controls Inducible Transcription of the Proinflammatory Cytokine IL-1α. *J Immunol* 2015; **195**: 1359–63.

157 Ranzani V, Rossetti G, Panzeri I, Arrigoni A, Bonnal RJP, Curti S *et al.* The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nat Immunol* 2015; **16**: 318–25.

158 Muljo SA, Zhu J, Zhao K. Expression and regulation of lincRNAs during T cell development and differentiation. *Nat Immunol* 2013; **14**: 1190–1198.

159 Zhang Y, Cao X. Long noncoding RNAs in innate immunity. *Cell Mol Immunol* 2016; **13**: 138–47.

160 Sigdel KR, Cheng A, Wang Y, Duan L, Zhang Y. The Emerging Functions of Long Noncoding RNA in Immune Cells: Autoimmune Diseases. *J Immunol Res* 2015; **2015**: 848790.

161 Imamura K, Akimitsu N. Long Non-Coding RNAs Involved in Immune Responses. *Front Immunol* 2014; **5**: 573.

162 Rapicavoli N a, Qu K, Zhang J, Mikhail M, Laberge R-M, Chang HY. A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *Elife* 2013; **2**: e00762.

163 Liu B, Sun L, Liu Q, Gong C, Yao Y, Lv X *et al.* A cytoplasmic NF-κB interacting long noncoding RNA blocks IκB phosphorylation and suppresses breast cancer metastasis. *Cancer Cell* 2015; **27**: 370–81.

164 Messemaker TC, Frank-Bertoncelj M, Marques RB, Adriaans a, Bakker a M, Daha N *et al.* A novel long non-coding RNA in the rheumatoid arthritis risk locus TRAF1-C5 influences C5 mRNA levels. *Genes Immun* 2016; **17**: 85–92.

165 Carpenter S, Aiello D, Atianand MK, Ricci EP, Gandhi P, Hall LL *et al.* A long noncoding RNA induced by TLRs mediates both activation and repression of immune response genes. *Science (80-)* 2014; **341**.

166 Mao A-P, Shen J, Zuo Z. Expression and regulation of long noncoding RNAs in TLR4 signaling in mouse macrophages. *BMC Genomics* 2015; **16**: 45.

167 IIott NE, Heward J a, Roux B, Tsitsiou E, Fenwick PS, Lenzi L *et al.* Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes. *Nat Commun* 2014; **5**: 3979.

168 Cui H, Xie N, Tan Z, Banerjee S, Thannickal VJ, Abraham E *et al.* The human long noncoding RNA, lnc-IL7R, regulates inflammatory response. *Eur J Immunol* 2014; : 1–36.

169 Li Z, Chao T-C, Chang K-Y, Lin N, Patil VS, Shimizu C *et al.* The long noncoding RNA THRIL regulates TNFα expression through its interaction with hnRNPL. Proc. Natl. Acad. Sci. U. S. A. 2014; **111**: 1002–7.

170 Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol* 2012; **30**: 1095–106.

171 Chen G, Qiu C, Zhang Q, Liu B, Cui Q. Genome-wide analysis of human SNPs at long intergenic noncoding RNAs. *Hum Mutat* 2013; **34**: 338–44.

172 Ding J, Eyre S, Worthington J. Genetics of RA susceptibility, what comes next? *RMD Open* 2015; **1**: e000028–e000028.

173 Pasmant E, Sabbagh A, Vidaud M, Bièche I. ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J* 2011; **25**: 444–8.

174 Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C *et al.* Population genomics of human gene expression. *Nat Genet* 2007; **39**: 1217–1224.

175 Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT. Mapping determinants of human gene expression by regional and genome-wide association. *Nature* 2005; **437**: 1365–9.

176 Hrdlickova B, Almeida RC De, Borek Z, Withoff S. Biochimica et Biophysica Acta Genetic variation in the non-coding genome : Involvement of micro-RNAs and long non-coding RNAs in disease ☆. *BBA - Mol Basis Dis* 2014; **1842**: 1910–1922.

177 Popadin K, Gutierrez-Arcelus M, Dermitzakis ET, Antonarakis SE. Genetic and epigenetic regulation of human lincRNA gene expression. *Am J Hum Genet* 2013; **93**: 1015–26.

178 Almlöf JC, Lundmark P, Lundmark A, Ge B, Pastinen T, Goodall AH *et al.* Single nucleotide polymorphisms with cis-regulatory effects on long non-coding transcripts in human primary monocytes. *PLoS One* 2014; **9**: e102612.

179 Mirza AH, Kaur S, Brorsson C a, Pociot F. Effects of GWAS-associated genetic variants on lncRNAs within IBD and T1D candidate loci. *PLoS One* 2014; **9**: e105723.

180 Spielmann M, Mundlos S. Looking beyond the genes: the role of non-coding variants in human disease. *Hum Mol Genet* 2016; **25**: R157–R165.

181 Plenge RM, Ph D, Seielstad M, Padyukov L, Lee T, Remmers EF *et al.* TRAF1–C5 as a Risk Locus for Rheumatoid Arthritis — A Genomewide Study. *NEJM* 2007; **357**: 1199–1209.

182 Kurreeman F a S, Goulielmos GN, Alizadeh BZ, Rueda B, Houwing-Duistermaat J, Sanchez E *et al.* The TRAF1-C5 region on chromosome 9q33 is associated with multiple autoimmune diseases. *Ann Rheum Dis* 2010; **69**: 696–9.

183 Messemaker T, Toes REM, Kurreeman F, Mikkers HMM. Comment on 'Functional Analysis of a Complement Polymorphism (rs17611) Associated with Rheumatoid Arthritis'. *J Immunol*; : 2.

184 Wang Z, Jinnin M, Nakamura K, Harada M, Kudo H, Nakayama W *et al.* Long non-coding RNA TSIX is upregulated in scleroderma dermal fibroblasts and controls collagen mRNA stabilization. *Exp Dermatol* 2016; **25**: 131–136.

Chapter 2

Inflammatory genes TNFα and IL6 display no signs of increased H3K4me3 in circulating monocytes from untreated rheumatoid arthritis patients

Genes Immun. 2017 Sep;18(3):191-196. doi: 10.1038/gene.2017.20.

Messemaker TC, Mikkers HM, Huizinga TW, Toes RE, van der Helm- van Mil AH, Kurreeman F

Abstract

Innate immune cells, such as monocytes, can adopt a long-lasting pro-inflammatory phenotype, a phenomenon called 'trained immunity'. In trained immunity, increased cytokine levels of genes, like interleukin (IL)-6 and tumor necrosis factor (TNF)-α, are observed, which are associated with increased histone 3 lysine 4 trimethylation (H3K4me3) in the promoter region. As systemic IL6 and TNFα levels are increased in rheumatoid arthritis (RA) patients and monocytes are known to be the primary producers of TNFα and IL6, we hypothesized that 'trained immunity' signals may be observed at these genes in monocytes from RA patients. CD14+ monocytes were isolated from untreated RA patients and paired age-matched healthy controls. H3K4me3, mRNA, protein and serum levels of IL6 and TNFα were evaluated by chromatin immunoprecipitation, reverse-transcription quantitative PCR and enzyme-linked immunosorbent assays. Despite elevated serum levels of TNFα and IL6 in the tested RA patients (P<0.05), *ex vivo* isolated monocytes displayed similar H3K4me3 levels to healthy controls in the promoter region of TNFα and IL6. Concordantly, mRNA and protein levels of IL6 and TNFα were similar before and after lipopolysaccharide stimulation between patients and controls. Together, with the current number of individuals tested we have not detected enhanced trained immunity signals in circulating monocytes from untreated RA patients, despite increased IL6 and TNFα serum levels.

Introduction

Rheumatoid arthritis (RA) is a chronic autoimmune disease mainly characterized by inflammation of the joints. Although some effective treatment of RA exists, many factors that lead to the development of the disease remain unknown. Both environmental and genetic factors have been shown to have a role in the onset of the disease.[1] For many diseases, including those of autoimmune nature, environmental triggers are thought to influence gene regulation via changes in the epigenetic landscape.[2,3] Epigenetic changes are defined by non-genetic changes in gene activity, which are influenced by various factors including the following: non-coding RNAs, DNA methylation and histone modifications.[4,5] In innate immune cells as well, environmental triggers were shown to induce long-lasting epigenetic changes described in a concept known as 'trained immunity'.[6] In the study from Quintin *et al.*,[6] monocytes were pre-exposed to β-glucan (a cell-

wall component of *Candida albicans*) for 24 h, which resulted in changes of the epigenetic landscape. One of the identified hallmarks of trained immunity was an induction of histone 3 lysine 4 trimethylation (H3K4me3) in the promotor region of immune genes like tumor necrosis factor (TNF)-α and interleukin (IL)-6, an indicative mark for active promoters.[6,7] As a result of increased H3K4me3 levels, a second immune stimulation of these pre-exposed monocytes resulted in an increased immune response shown initially by elevated IL6 and TNFα levels.[6] These results indicated that innate immune cells like monocytes can acquire long-lasting alterations in the epigenetic landscape resulting in the enhanced production of cytokines.

In RA, TNFα and IL6 are elevated in the serum and synovium of patients compared with healthy individuals.[8] Moreover, both anti-TNFα treatment and anti-IL6 treatment are currently successful therapies against RA, suggesting the importance of both cytokines in disease pathogenesis.[9] As monocytes can produce high levels of TNFα and IL6, we hypothesized that changes in H3K4me3 of monocytes might exist in early untreated RA patients, thereby contributing to enhanced cytokine levels and thus to the disease pathogenesis of RA.[10] To our knowledge, this is the first report measuring H3K4me3 levels in the promoter region of two RA-relevant cytokines in primary immune cells.

Results

Detection of differences in H3K4me3 levels in the promoter of IL6 and TNFα in human monocytes

In order to test our hypothesis and to show that in our hands monocytes can obtain a trained phenotype, we repeated the previously published trained immunity model as described by Quentin *et al.*[6] Monocytes from healthy donors were pre-incubated with β-glucan for 24 h followed by a 6-day resting/refreshing period. After 7 days, H3K4me3, protein and RNA levels of TNFα and IL6 from these monocytes were compared with monocytes, which were not exposed to β-glucan from the same donor. We observe similar effects with a ~2-fold increase in H3K4me3 levels at the TNFα and IL6 promoter regions with a concurrent increase in RNA and cytokine levels (Supplementary Figure S1). We next calculated how many RA patients and matched healthy controls we would require to detect

similar effect sizes. A total sample size of four is required, two RA patients and two healthy controls, providing 80% to detect a significant difference at $P<0.05$.

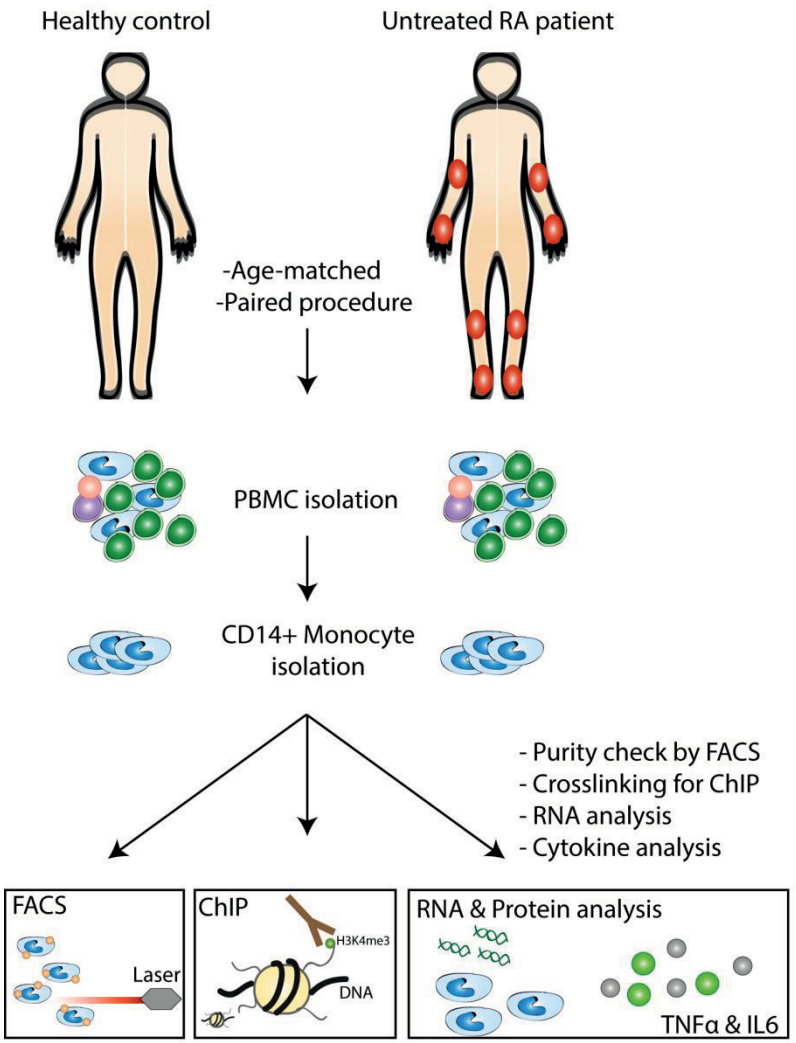

Figure 1. Flowchart of technical procedure isolating monocytes from healthy controls and untreated RA patients. Blood was obtained from a newly diagnosed untreated RA patient and a healthy control on the same day and used for PBMCs isolation. After a paired PBMC isolation, CD14+ monocytes were isolated by magnetic beads and used for (i) purity check by fluorescence-activated cell sorting, (ii) DNA-Histone crosslinking and stored at −80 °C for ChIP experiments or (iii) RNA and protein analysis by culturing the isolated monocytes in presence of RPMI (unstimulated) or LPS.

Similar H3K4me3 levels on the IL6 and TNFα promoter in monocytes from RA patients and healthy controls

To minimize sources of variation between patients and controls, we simultaneously collected blood and serum from RA patients and age-matched healthy individuals (Figure 1 and Supplementary Figure S2). RA patients display significantly higher IL6 and TNFα levels in serum as compared with controls, indicating that *in vivo* differences in cytokine levels are present (Supplementary Figure S3). In order to measure H3K4me3 levels, CD14+ monocytes were directly isolated and H3K4me3 levels of TNFα and IL6 were simultaneously measured. However, H3K4me3 levels on the promoter region of TNFα and IL6 were similar in untreated RA patients and healthy controls (Figures 2a and b, left panel, *P*=0.61 for both genes). The observed content of H3K4me3 from our experiments coincide with previously measured H3K4me3 chromatin immunoprecipitation (ChIP)-sequencing levels in healthy individuals derived from publicly available data from ENCODE/broad institute (Figures 2a and b, right panel). H3K4me3 levels of housekeeping gene RPL5 were similar between untreated RA patients and healthy controls, and correcting for RPL5 as an additional internal control did not alter the H3K4me3 levels between healthy controls and untreated RA patients (Supplementary Figure S4A and B). Finally, Insulin, which is not expressed by monocytes, was used as a negative control and did not display H3K4me3 levels in monocytes as expected (Supplementary Figure S4C).

Similar IL6 and TNFα mRNA and protein levels, and responsiveness of monocytes from RA patients and healthy controls

To confirm our histone trimethylation results and further confirm that both monocytes from RA patients and healthy controls respond similarly, we evaluated RNA levels of TNFα and IL6 before and after lipopolysaccharide (LPS) stimulation. In unstimulated monocytes we observed low levels of TNFα and IL6 that were similar in controls and RA patients (Supplementary Figure S5). Upon stimulation, a strong induction of expression was observed in both healthy controls and RA patients for TNFα and IL6 (Figures 3a and b). Again, similar RNA levels between both groups were observed for IL6 (RA: 42637±19208 vs controls: 47996±38235; *P*=0.70) and TNFα (RA: 78.5±20.9, controls: 83.5±42.3; *P*=0.9). In addition, we investigated cytokine levels in the supernatant of unstimulated and LPS-stimulated monocytes. Similar protein levels were observed in the

supernatant of monocyte cultures obtained from the two groups (Figures 3c and d), although TNFα and IL6 levels seemed slightly elevated in RA patients after LPS-stimulation (TNFα concentration: RA: 4.3 ng ml^{-1} ±2.9 vs controls: 2.8 ng ml^{-1} ±2.6; *P*=0.40); IL6 concentrations: RA: 67.2 µg ml^{-1} ±14.6 vs controls: 50.8 µg ml^{-1} ±17.3; *P*=0.19). In addition, dividing patients on basis of autoantibody status (anti-citrullinated protein antibody positivity) revealed neither differences in IL6 and TNFα production (Supplementary Figure S6). Together, these findings suggest that there are no substantial differences regarding H3K4me3, mRNA levels or cytokine levels of IL6 and TNFα between circulating monocytes from RA patients and healthy controls in this study.

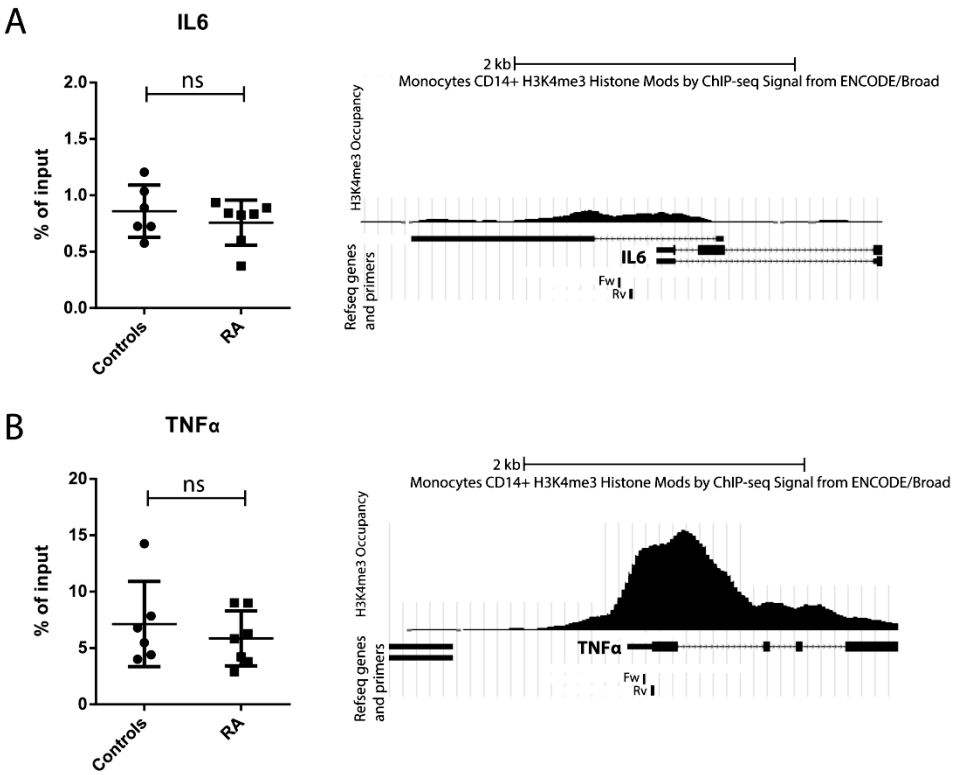

Figure 2. H3K4me3 levels in monocytes from healthy controls and untreated RA patients. (a) H3K4me3 levels of the IL6 promoter in healthy controls (controls) and untreated RA patients (RA) (top). The y axis represents the percentage of precipitated H3K4me3 normalized for DNA input. H3K4me3 occupancy of the IL6 promoter in healthy individuals measured by ChIP-sequencing (ChIP-seq) from ENCODE/Broad institute (bottom). (b) H3K4me3 levels of the TNFα promoter in healthy controls (controls) and untreated RA patients (RA) (top). H3K4me3 occupancy of the TNFα promoter in healthy individuals measured by ChIP-seq from ENCODE/Broad institute (bottom). Mann–Whitney *U*-test was used to test for significant differences.

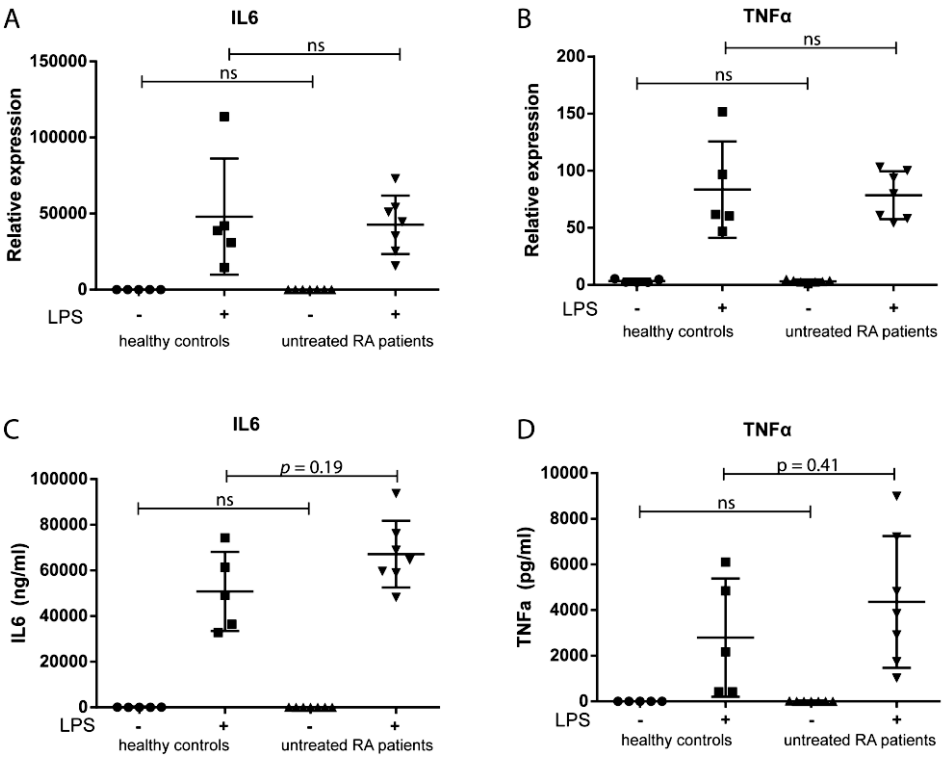

Figure 3. RNA and protein levels of IL6 and TNFα in monocytes from healthy controls and RA patients. RNA levels of IL6 (a) and TNFα (b) were determined for unstimulated (−) and LPS-stimulated (+) monocytes isolated from healthy controls and untreated RA patients. RNA levels were normalized for *GAPDH* and *B2M* as housekeeping genes. The sample with the lowest expression was set to 1. Protein concentrations of IL6 (c) and TNFα (d) were measured in the supernatant of unstimulated (−) and LPS-stimulated (+) monocytes isolated from healthy controls and untreated RA patients. The concentration of LPS added was 10 ng ml^{-1}. Mann–Whitney *U*-test was used to test for significant differences.

Discussion

In this study we aimed to measure H3K4me3 levels in the promotor regions of TNFα and IL6, in *ex vivo* primary monocytes from RA patients and healthy individuals. TNFα and IL6 are important cytokines for the pathogenesis of RA and can both be found in large abundances in the joints of RA patients.[11,12] IL6 can activate several cell types and is involved in the maturation of B cells and thus the production of autoantibodies.[13] TNFα is a potent inducer of other inflammatory

genes and inhibition of either TNFα or IL6 have both been shown to be effective for treatment of RA.[14,15] Monocytes are the primary producers of IL6 and TNFα and 'trained' monocytes have been shown to have enhanced production of both cytokines. In order to test whether monocytes from RA patients display different epigenetic patterns at these two genes, a rigorous protocol of collection of blood and isolation of monocytes from both RA patients and paired age-matched healthy controls was implemented. To our knowledge, this is the first time that a study addressing epigenetic changes in *ex vivo* primary monocytes from early untreated RA patients has been performed.

Epigenetic modifications have a crucial role in regulating gene expression.[16] The role of DNA methylation and histone modifications have both been studied in relation to RA.[17] In synovial fibroblast, structural changes in DNA methylation correlate with fibrotic development in the joint of RA patients.[18] In several immune cells as well, changes were observed in the DNA methylation of relevant immune-related genes, for example, hypomethylation of IL6 in B cells of RA patients[19] or hypomethylation of CD40LG in CD4+ T cells.[20] Furthermore, expression levels of various histone-modifying enzymes have been investigated in RA. EZH2 (Enhancer of Zeste Homolog 2), a histone methyl transferase was found elevated in synovial fibroblast from RA patients.[21] However, studies evaluating changes in histone methylation in patients with autoimmune diseases are scarce. In this study, we evaluated for the first time H3K4me3 levels on the promoter of two important cytokine genes in monocytes directly isolated from RA patients.

Although many types of histone modifications are known, H3K4me3 was shown to correlate with higher TNFα and IL6 levels in β-glucan-trained monocytes and was therefore used as a marker for trained immunity. Here we report that although we can reliably measure twofold differences in H3K4me3 levels between β-glucan-trained and untrained monocytes, our study was sufficiently powered to detect effect size differences as low as 1.8; however, such epigenetic differences on the H3K4me3 mark between RA patients and healthy controls were not observed. This finding was confirmed by RNA and protein levels of IL6 and TNFα, which had similar levels in RA patients and healthy controls after LPS stimulation. These results indicate that RA monocytes are likely not in a continuous pro-inflammatory ON-state at the IL6 and TNFα genes on basis of the H3K4me3 levels and LPS responsiveness. Although we demonstrate that *in vitro*

trained monocytes have similar time-response dynamics in their cytokine production as untrained monocytes (Supplementary Figure S7), we cannot exclude that RA-specific differences in cytokine dynamics exist. It is probable that RA monocytes produce higher TNFα and IL6 levels *in vivo* via other mechanisms as the possibility exists that RA monocytes will stay activated longer compared to healthy controls and such effects will be missed in our study due to the chosen time points and LPS concentration. Moreover, in our study, monocytes were obtained from peripheral blood and we cannot exclude that locally activated monocytes in the joints of RA patients show changes in their epigenetic landscape. Besides monocytes other RA-relevant cells types are capable of producing TNFα and/or IL6. These cell types include T cells, B cells, mast cells, synovial fibroblasts and osteoblasts, and may also contribute to the induced TNFα and IL6 levels in the serum of patients.[22] It would be interesting to investigate whether these cell types can also obtain a trained epigenetic landscape and whether this contributes to development of RA.

Besides changes in H3K4me3 upon immune stimulation, other histone marks such as H3K4me1 and H3K27Ac have more recently been described to shape the trained immunity epigenetic landscape.[23] In line with these findings, latent enhancers also acquire changes in H3K4me1, H3K4me3, H3K27Ac and PU.1 binding upon immune triggering.[24,25] These studies illustrate the complexity of the epigenetic landscape that is formed upon different immune stimuli and suggests that further in-depth studies may be required. Nonetheless, epigenetic changes are currently being used for diagnostics and therapy in other diseases (reviewed in [ref. 26]). In RA as well, an inhibitor of histone-modifying enzyme was suggested as an inflammatory suppressor.[27] Gillespie *et al.*[27] showed that RA peripheral blood mononuclear cells (PBMCs) possess increased histone deacetylase activity and inhibition of this class of enzymes could reduce the production of both TNFα and IL6 in a compound and cell-type-dependent manner.[27] However, PBMCs are a collection of diverse cell types and results obtained represent an average across all cells. These data highlight the complexity and variety of factors that can influence the epigenetic landscape. Investigating the epigenetic landscape in disease pathology is thus influenced by these factors and the outcome of results may rely on (i) the histone modification that is being studied, (ii) the genes investigated and (iii) the choice of cell type. Currently, our results indicate that large differences between circulating

monocytes isolated from RA patients and healthy individuals are absent in context of TNFα and IL6 levels. However, we cannot exclude that more subtle differences may be detected in larger sample sizes. In conclusion, with the current number of individuals tested we have not detected evidence that H3K4me3 levels at two immune relevant genes IL6 and TNFα are enhanced in RA. Further studies addressing genome-wide patterns of epigenetic changes in larger numbers of individuals may yield additional insight into whether epigenetic changes in monocytes may be relevant for the pathogenesis of RA.

Materials and methods

Patients

Patients fulfilling the 1987 criteria for RA[28] recruited from the rheumatologic outpatient clinic of the Leiden University Medical Center were included in this study. Patients were studied at the time of diagnosis and before disease-modifying antirheumatic drugs (including no corticosteroids) were started. Age-matched healthy controls were also included, and were preferably healthy partners or family member of the patients. This study was approved by the ethical committee of the Leiden University Medical Center and all patients and healthy donors provided written informed consent. Serum from patients and healthy controls was obtained by incubation of blood at room temperature for 30 min followed by centrifugation at 2000 *g* for 10 min. Supernatant (serum) was transferred and stored at −80 °C until Illuminex measurements. PBMCs isolations were performed from an untreated RA patient and a matched healthy control and were performed on the same day in a paired manner, with the exception one RA patient, which did not have a paired control. PBMC isolations were performed from eight untreated RA patients (seven females, one male mean±s.d. age 59.3±11.2 years) and seven healthy controls (five females, two males; age 51.6±10.9 years). Four patients were positive for anti-citrullinated protein antibodies and five patients were positive for rheumatoid factor. More detailed characteristics can be found in Supplementary Table S1.

Cell culture

PBMCs from healthy controls and RA patients were isolated on a Ficoll gradient (Pharmacy LUMC, Leiden, The Netherlands). CD14+ monocytes were isolated using magnetic beads and cultured as previously described.[29] Cells were either

crosslinked for ChIP experiments and/or cultured for RNA and protein measurements. For the ChIP experiment, seven RA patients and six healthy controls were used. For seven RA patients and five healthy controls, enough cells were obtained from CD14+ isolation and additional RNA and protein analysis was performed. RNA was isolated 4 h after stimulation with either LPS or RPMI (unstimulated). Supernatant was collected 20 h after stimulation with LPS or from unstimulated monocytes and used for enzyme-linked immunosorbent assay to measure the level of secreted cytokines. β-Glucan stimulation was performed as described by Quintin *et al.*[6] In short, CD14+ isolated monocytes were once pre-incubated for 24 h with either β-glucan (Invivogen, San Diego, CA, USA) from *Alcaligenes faecalis* (10 μg ml^{-1}) or RPMI medium (unstimulated). After 24 h, β-glucan was removed by refreshing with new medium without β-glucan and was continued to be refreshed every 2 days over a 6-day period. After 7 days (1 day β-glucan, 6 day washing/refreshing), monocytes were either used for chromatin immune precipitation or stimulated with LPS (10 ng ml^{-1}) and RPMI (unstimulated) for RNA and protein analysis. Viability of monocytes after 7 days of culturing was monitored by a LUNA-II automated cell counter (Logos Biosystems, Villeneuve d'Ascq, France).

Chromatin immunoprecipitation

For ChIP, 1.5 × 10^6 monocytes were crosslinked in 1% formaldehyde (Mallinckrodt-Baker, Dublin, Ireland) for 10 min, treated with 1/10 volume of 1.25 M glycine (J.T.Baker, Thermo Fisher, Waltham, MA, USA) for 3 min, washed with ice-cold phosphate-buffered saline and stored in −80 °C after snap freezing. Crosslinked monocytes were lysed using 1% SDS and 20 mM HEPES, and sonicated using a Biorupter sonication device (Diagnode, Seraing, Belgium). Chromatin sonication (DNA fragments between 150–600 bp) was validated by agarose gel electrophoresis. Sonicated chromatin was divided into two fractions and used as either input material or for overnight H3K4me3 precipitation using protein G dynabeads (Life Technologies, Oslo, Norway) and polyclonal H3K4me3 antibody (Catalog number C15410003-010, Diagnode). DNA was isolated by de-crosslinking input DNA and H3K4me3 precipitated DNA in elution buffer (1% SDS, 0.1 M NaHCO3, 0.19 M NaCl, 2 units Prot K (Thermo Scientific)) at 65 °C for 4 h followed by phenol/chloroform DNA precipitation. Quantification of immunoprecipitated DNA was performed by quantitative PCR using primers provided in Supplementary Table S2. Primers were located in the promotor

regions of IL6, TNFα and RPL5 as shown by H3K4me3 containing regions in monocytes via publicly available ChIP-sequencing data obtained from ENCODE/Broad institute (GEO accession: GSM1003536). Insulin, which is not expressed by monocytes, was used as negative control.

RNA isolation, cDNA synthesis and expression analysis

RNA isolation and cDNA synthesis was performed as previously described.[29] Two housekeeping genes were used for normalization: glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) and β-2 microglobulin (*B2M*). Relative expression was calculated using the ΔΔCT method.[30] The sample with the lowest expression was set to 1. Primer sequences are listed in Supplementary Table S2. Before quantitative PCR, minus RT samples were checked for genomic DNA contamination.

Cytokine production

TNFα and IL6 levels in the medium of cultured monocytes were measured using human TNF ELISA kit (BD Bioscience, San Jose, CA, USA) or human IL6 DuoSet ELISA (R&D System, Abingdon, UK), respectively. TNFα and IL6 in serum were determined using the Milliplex Human Adipocyte panel (Millipore, Amsterdam, The Netherlands) and measured on the Bio-Plex array reader and Bio-Plex software in accordance with the manufacturer's instructions.

Statistical analysis

Data are presented as means±s.d. Statistical analysis was performed using GraphPad (GraphPad, San Diego, CA, USA). Mann–Whitney *U*-test was performed to determine significant differences. Differences with *P*-values <0.05 were considered statistically significant. Sample size calculation was performed based on the observed effect size in β-glucan-trained monocytes using G power v3.1 (University of Düsseldorf, Germany).

Supplementary information

Supplementary information is available online on the Genes and Immunity website (http://www.nature.com/gene): Supplementary Figure S1-S7 and Supplementary Table S1 and S2.

Acknowledgements

We are grateful for the RA patients and the healthy volunteers for donating blood. This work was supported by the Dutch Arthritis Foundation, The Netherlands. FK has support from UNESCO-L'Oreal for Women in Science Fellowship, Marie Curie FP7 Outgoing Fellowship and an LUMC Research Fellowship. The work of VdH-vM was supported by a Vidi-grant of the Dutch Organisation for Scientific Research.

References

1. Messemaker TC, Huizinga TW, Kurreeman F. Immunogenetics of rheumatoid arthritis: Understanding functional implications. J Autoimmun 2015; 64: 7–14.
2. Viatte S, Plant D, Raychaudhuri S. Genetics and epigenetics of rheumatoid arthritis. Nat Rev Rheumatol 2013; 9: 141–153.
3. Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. Nat Rev Genet 2007; 8: 253–262.
4. Ballestar E. Epigenetics lessons from twins: prospects for autoimmune disease. Clin Rev Allergy Immunol 2010; 39: 30–41.
5. Peschansky VJ, Wahlestedt C. Non-coding RNAs as direct and indirect modulators of epigenetic regulation. Epigenetics 2014; 9: 3–12.
6. Quintin J, Saeed S, Martens JHa, Giamarellos-Bourboulis EJ, Ifrim DC, Logie C *et al. Candida albicans* infection affords protection against reinfection via functional reprogramming of monocytes. Cell Host Microbe 2012; 12: 223–232.
7. Netea MG, Joosten LaB, Latz E, Mills KHG, Natoli G, Stunnenberg HG *et al.* Trained immunity: a program of innate immune memory in health and disease. Science 2016; 352: aaf1098.
8. Wei S-T, Sun Y-H, Zong S-H, Xiang Y-B. Serum levels of IL-6 and TNF-α may correlate with activity and severity of rheumatoid arthritis. Med Sci Monit 2015; 21: 4030–4038.
9. Xu S. TNF inhibitor therapy for rheumatoid arthritis. Biomed Rep 2012; 1: 177–184.
10. Agarwal S, Piesco NP, Johnst LP, Riccellit AE. Differential expression of IL-113, TNF-ox, IL-6, and IL-8 in human monocytes in response to lipopolysaccharides from different microbes. J Dent Res 2016; 74: 1057–1065.
11 .Houssiau FA, Devogelaer JP, J Van Damme, CN de Deuxchaisnes, Van Snick J. Interleukin-6 in synovial fluid and serum of patients with rheumatoid arthritis and other inflammatory arthritides. Arthritis Rheum 1988; 31: 784–788.
12. Tetta C, Camussi G, Modena V, Vittorio CDi, Baglioni C. Tumour necrosis factor in serum and synovial fluid of patients with active and severe rheumatoid arthritis. Ann Rheum Dis 1990; 49: 665–667.
13. Srirangan S, Choy EH. The role of interleukin 6 in the pathophysiology of rheumatoid arthritis. Ther Adv Musculoskelet Dis 2010; 2: 247–256.
14. Woodrick R, Ruderman EM. Anti-interleukin-6 therapy in rheumatoid arthritis. Bull NYU Hosp Jt Dis 2010: 68 211–217.
15. Vasanthi P, Nalini G, Rajasekhar G. Role of tumor necrosis factor-alpha in rheumatoid arthritis : a review. Int J Rheum Dis 2007; 10: 270–274.
16. Greer EL, Shi Y. Histone methylation: a dynamic mark in health, disease and inheritance. Nat Rev Genet 2012; 13: 343–357.
17. Klein K, Ospelt C, Gay S. Epigenetic contributions in the development of rheumatoid arthritis. Arthritis Res Ther 2012; 14: 227.
18. Karouzakis E, Gay RE, Michel Ba, Gay S, Neidhart M. DNA hypomethylation in rheumatoid arthritis synovial fibroblasts. Arthritis Rheum 2009; 60: 3613–3622.

19. Nile CJ, Read RC, Akil M, Duff GW, Wilson AG. Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis. Arthritis Rheum 2008; 58: 2686–2693.
20. Liao J, Liang G, Xie S, Zhao H, Zuo X, Li F *et al.* CD40L demethylation in CD4(+) T cells from women with rheumatoid arthritis. Clin Immunol 2012; 145: 13–18.
21. Trenkmann M, Brock M, Gay RE, Kolling C, Speich R, Michel BA *et al.* Expression and function of EZH2 in synovial fibroblasts: epigenetic repression of the Wnt inhibitor SFRP1 in rheumatoid arthritis. Ann Rheum Dis 2011; 70: 1482–1488.
22. McInnes IB, Schett G. Cytokines in the pathogenesis of rheumatoid arthritis. Nat Rev Immunol 2007; 7: 429–442.
23 .Saeed S, Quintin J, Kerstens HHD, Rao Na, Aghajanirefah A, Matarese F *et al.* Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. Science 2014; 345: 1251086.
24. Kaikkonen MU, Spann NJ, Heinz S, Romanoski CE, Allison Ka, Stender JD *et al.* Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Mol Cell 2013; 51: 310–325.
25. Ostuni R, Piccolo V, Barozzi I, Polletti S, Termanini A, Bonifacio S *et al.* Latent enhancers activated by stimulation in differentiated cells. Cell 2013; 152: 157–171.
26. Zhang Z, Zhang R. Epigenetics in autoimmune diseases: pathogenesis and prospects for therapy. Autoimmun Rev 2015; 14: 854–863.
27. Gillespie J, Savic S, Wong C, Hempshall A, Inman M, Emery P *et al.* Histone deacetylases are dysregulated in rheumatoid arthritis and a novel histone deacetylase 3-selective inhibitor reduces interleukin-6 production by peripheral blood mononuclear cells from rheumatoid arthritis patients. Arthritis Rheum 2012; 64: 418–422.
28. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS *et al.* The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988; 31: 315–324.
29. Messemaker TC, Frank-Bertoncelj M, Marques RB, Adriaans A, Bakker AM, Daha N *et al.* A novel long non-coding RNA in the rheumatoid arthritis risk locus TRAF1-C5 influences C5 mRNA levels. Genes Immun 2016; 17: 85–92.
30. .Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001; 25: 402–408.

Chapter 3

Immunogenetics of rheumatoid arthritis: Understanding functional implications

J Autoimmun. 2015 Nov;64:74-81. doi: 10.1016/j.jaut.2015.07.007.

Messemaker TC, Huizinga TW, Kurreeman F

Abstract

The last decade has seen a dramatic technological revolution. The characterisation of the majority of the common variations in our genetic code in 2003 precipitated the discovery of the genetic risk factors predisposing to Rheumatoid Arthritis development and progression. Prior to 2007, only a handful of genetic risk factors had been identified, HLA, PTPN22 and CTLA4. Since then, over 100 genetic risk loci have been described, with the prediction that an ever-increasing number of risk alleles with consistently decreasing effect sizes will be discovered in the years to come. Each risk locus harbours multiple candidate genes and the proof of causality of each of these candidates is as yet unknown. An enrichment of these RA-associated genes is found in three pathways: T-cell receptor signalling, JAK-STAT signalling and the NF-κB signalling cascade, and currently drugs targeting these pathways are available for the treatment of RA. However, the role that RA-associated genes have in these pathways and how they contribute to disease is not always clear. Major efforts in understanding the contribution of genetic risk factors are currently under way with studies querying the role of genetic variation in gene expression of coding and non-coding genes, epigenetic marks and other regulatory mechanisms yielding ever more valuable insights into mechanisms of disease. Recent work has suggested a possible enrichment of non-coding RNAs as well as super-enhancers in RA genetic loci indicating possible new insights into disease mechanism. This review brings together these emerging genetic data with an emphasis on the immunogenetic links these findings have provided and what we expect the future will bring.

1. Introduction

Rheumatoid arthritis (RA) is a heterogeneous chronic (auto)immune disease associated with significant morbidity and reduced life expectancy. Global prevalence of RA has been estimated to be around 0.2–0.5% on average, with a large variation across regions [1,2]. The highest prevalence has been detected in Europe and North America with lower prevalence in Africa and Southeast Asia. In general, there is a two-fold higher occurrence in females than in males. Given the common prevalence and the lack of a cure for RA, the socio-economic burden remains large and is predicted to rise with an increasingly ageing population [3]. Rheumatoid Arthritis is characterized by chronic inflammation and destruction of the synovial joints leading to progressive joint damage and disability.

Autoimmunity, identified by the production of auto-antibodies such as rheumatoid factor (RF) or anti-citrullinated protein antibodies (ACPA) precedes the clinically detectable onset of inflammatory arthritis and can last for years (these aspects have been reviewed elsewhere in detail) [4]. Individuals who harbour autoantibodies tend to have a more severe disease course and respond differently to treatment as compared to those who do not [5]. Interestingly though, at the time of diagnosis, no difference has so far been detected in clinical presentation of autoantibody positive patients versus autoantibody negative patients.

Both genetic and environmental factors are thought to play a role in disease development and disease progression. The heritable component of RA is evident from the 15% concordance rate observed in monozygotic twin pairs and increased familial clustering [6,7]. Heritability estimates of autoantibody positive individuals are similar to autoantibody negative individuals (~40–50%) indicating a significant contribution of genetic factors to both subgroups [8].

Identifying genetic factors has largely been hampered by the existence of genetic heterogeneity, low penetrance of individual disease alleles and the potential for gene–gene/gene–environment interactions. Nevertheless, candidate gene studies but to a larger extent genome-wide association studies querying ~10 million variants in the human genome in ~100,000 individuals have led to the identification of >100 loci that are associated with RA [9]. These loci individually confer only modest effects decreasing their potential utility in the clinic as a prediction tool but do provide important insights into relevant pathways involved in the disease process. It is important to note that the majority of association studies have been performed in individuals of European ancestry and patients who harbour autoantibodies and our review mainly discusses the immunogenetic pathways from this relatively homogenous group of patients.

2. Immunogenetics of the HLA association with RA

The main genetic region linked to RA over thirty years ago, before the advent of genome wide association studies (GWAS), is the HLA region which is encoded by the major histocompatibility complex (MHC). The MHC locus spans approximately 4 Mb and contains approximately 250 genes, of which ~60% have immune-

related functions. The region is characterised by extended and complex linkage disequilibrium patterns that have made it notoriously difficult to pinpoint the causal gene(s) in the region. The initial association between HLA and RA was made in 1976 with the observation of an overrepresentation of HLA-DR4 in mixed lymphocyte cultures of RA patients [10]. Other HLA-DR molecules associate with RA defined by a common amino acid sequence in the HLA-DRB1 chain, termed the HLA shared epitope (HLA-SE) [11].

Over the last few years with the advent of GWAS to measure millions of variants along with the possibility to deeply sequence our genome, significant progress has been made to assess the association of the HLA region to autoantibody positive RA. More precisely, amino acid positions 11, 13, 71 and 74 at the HLA-DRB1 chain as well as position 9 of HLA-B and position 9 of HLA-DPB1 have now been identified as being the most statistically significant associations [12]. These positions are located within the antigen-binding groove to the HLA molecule further supporting the role of T cells in RA. Similar associations at the HLA region do exist in African Americans and East Asians, indicating possible shared mechanisms in different ethnic groups although more well powered studies need to be performed to dissect the overlap and differences at the HLA alleles [13]. In contrast, clearly distinct association signals (e.g. HLA-DR3) have been observed at the HLA locus in ACPA-negative individuals of European descent, shedding light on different genetic predispositions in the two disease subgroups [14–16]. Other HLA haplotypes such as HLA-DRB1*13 carrying the five amino acid sequence DERAA at positions 70–74 protect against development of RA [17,18]. These protective effects are confined to ACPA + patients indicating a possible overlap in pathways mediating risk and protection. While methods have been developed to allow the simultaneous query of hundreds of thousands of samples (which represents significant progress), very few new insights have been generated in elucidating the functional mechanisms underlying the HLA association with RA.

Importantly, a recent study has shown that HLA-DQ molecules, which are in full linkage disequilibrium with HLA-SE alleles, are able to efficiently present DERAA epitopes derived from microorganisms as well as from a self-protein known as vinculin [19]. DERAA-directed T cells can provide help to B cells ultimately leading to ACPA production. Individuals who carry HLA-DR13 tend have an decreased number of DERAA-directed T cells likely due to negative selection in the thymus

providing some additional clue of the role of the HLA locus in disease development [19]. Such studies provide an exciting avenue for future research on how HLA-peptide interactions shape the T-cell repertoire. Interestingly it has also been described that non-inherited maternal antigens expressed by the mother but not by the child are also able to provide protection [20]. This observation holds the promise that exposure to external antigens such as DERAA derived from micro-organisms in individuals with distinct genetic background may lead to protection from developing RA. The future will learn whether this pathway can be exploited to prevent RA in high risk individuals.

3. Non-HLA genetic risk factors

Prior to 2007 only a handful of genes outside of the HLA region had been identified including PTPN22 and CTLA4 in Europeans and PADI4 in Asians. In 2007, the TRAF1-C5 locus was concurrently discovered by a candidate gene approach as well the first genome-wide association study in RA. Research in this area, propelled by unparallelled efforts to (i) sequence human genomes (since 2001) [21], (ii) to characterise the most common genetic variations in human populations (HapMap www.HapMap.org [22,23], 1000 genomes project [24], since 2003), (iii) to reliably impute unmeasured genetic variation through robust statistical methods [25] (iv) to define more homogenous groups of patients (e.g through ACPA positivity), has seen a tremendous increase in the number of genetic regions associated with the susceptibility to RA. 101 loci have now been identified either at genome-wide significant thresholds and/or with evidence from replication studies [9,26–42].

The latest major study encompassed a combined analysis of 100,000 individuals of European and Asian descent with the query of ~10 million single nucleotide variants across the human genome. HLA remains the strongest association to disease with an odds ratio of 2–3 with the second strongest genetic risk being conferred by PTPN22 (OR 1.8) (Fig. 1). The remainder of genetic risk factors have modest effect sizes (<1.2) with a prediction of ever-decreasing odds ratios paired with an ever increasing number of risk alleles which will be discovered as sample sizes increase [43]. HLA explains the majority of the genetic risk ~13% with an additional 5% of the genetic risk being explained by an additional 100 loci discovered to date [9,12].

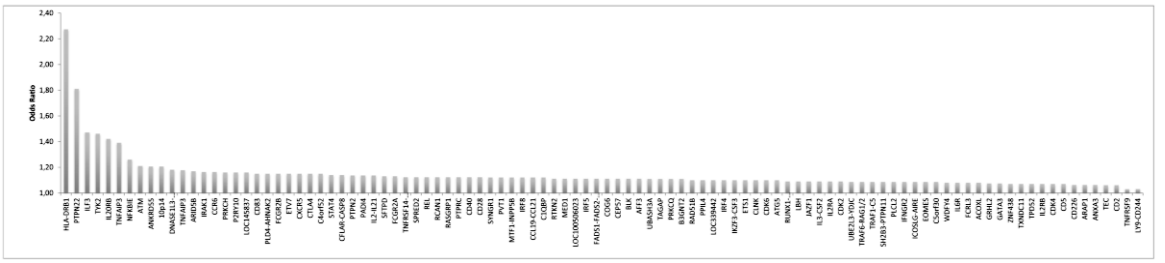

Fig 1. 101 Genetic risk loci predisposing to the development of Rheumatoid Arthritis. Susceptibility loci are ranked by effect size (Odds ratio, y axis) as observed in a meta-analysis performed in 100,000 individuals of European and Asian Ancestry. Each locus is identified by the most likely biological candidate in the associated region as provided by Okada *et al*. [9].

4. Functional implications of genetic risk loci identified to date

There are major challenges to understanding how genetic variation is involved in disease development. An association with a genetic variant does not directly lead to either a causal variant or a causal gene, making the task of translating the functional consequences of genetic variation in diseases where ORs are very low rather challenging. Importantly, the fact that parts of our human genome are inherited in blocks (linkage disequilibrium, LD) [22,23] makes the identification of causal genes and causal variants complicated. The approach currently employed in the identification of causal variants is (i) identify all variants that are (highly) linked to the best signal of association (ii) determine what functional consequences these variants may have (ie are they located in an exon, intron or intergenic region and do they result in a change in protein structure, function or expression). In the end, empirical experimental evidence is required to determine the effects of causal variants and genes and their contribution to the pathogenesis of disease.

In order to understand the functional consequences of genetic findings, there are a few crucial questions. (i) Which SNP will be chosen (ii) what is the endpoint to be measured (for example which gene expression should be measured) and finally (iii) in which cell-type should it be measured and should the cell be activated in order to detect putative differences? Despite the simplicity of the questions, there is complexity at all levels all of which are being addressed by the research community and which will in the end hopefully help us in understanding disease processes better.

Among the 100 non-HLA genetic loci identified to date (Supplementary Table 1), associated regions contain on average ~4 putative candidate genes (in total 377 genes across 100 loci). Within these regions, genetic variants in coding genes that lead to changes in amino acids resulting in dysfunctional proteins are very few (19 out of the 377 genes observed from the latest GWAS analyses, Table 1). Among the genes that have been reported to harbour at least one genetic variant with missense SNP(s) in high LD (R2 > 0.8), only a few have been linked to functional changes in candidate genes (Table 1).

One example of a functional missense variant was identified in PTPN22. The variant (rs2476601) is located in exon 14 and induces an amino acid substitution from an arginine to a trypthophan. This R620W conversion is located in protein motif thought to be required for protein–protein interaction [44]. Elegant studies have shown that the mutant (risk) allele results in decreased TCR and BCR signalling in lymphocytes [45] (reviewed recently by Rawlings *et al*. and by Burn *et al*. [46,47]). Interestingly, in mice, the homologous R620W variant in PTPN22 known as R619W located in Ptpn22 reduced protein levels of Ptpn22 and was shown to manifest in thymic and splenic enlargements [44,48]. This reduced Ptpn22 protein expression has been shown to diminish its inhibitory effect on T- and B-cell activation leading to an increased number of T cells, and an enhanced T-cell, dendritic-cell and B-cell activation. Furthermore the R-to-W conversion seems to increase the resistance of B cells to apoptosis and expands the pool of transitional and auto-reactive B cells. More recently, PTPN22 has been reported to interact with PADI4 (Peptidyl Arginine Deaminase 4), also a risk factor for RA in Asians and European [9]. PADI4 is involved in regulating the citrullination process through which ACPA may be generated. Polymorphisms in the PADI4 haplotype have been shown to affect the mRNA stability of the gene [49]. This is an example of new insight gained into pathogenesis generated by genetic findings. It is known that tolerance to citrullinated antigens is broken in RA and genetic studies have involved HLA in this process [50]. It is not known that the amount of antigen is involved in breaking of tolerance and if the stability of mRNA of PADI4 translates in different levels of the enzyme and subsequent amount of antigen, this is a new hypothesis put forward by genetic data. Interestingly, PTPN22 was recently shown to physically interact with PADI4. Deficiency of PTPN22 led to enhanced protein citrullination and formation of neutrophil extracellular traps (a mechanism in place to combat pathogens [51]). The data suggests that the

PTPN22 risk allele disrupts the interaction of PTPN22 with PADI4 leading to hypercitrullination in peripheral blood mononuclear cells [52]. Although these experiments require independent replication, they do provide novel insights into previously unknown mechanisms that could be at play in disease process and highlight the need to look beyond known interactions between proteins.

Table 1. GWAS candidate genes harbouring missense SNPs (Adapted from Okada *et al*.). Missense SNPs in high linkage disequilibrium with the lead SNP (Supplementary Table 1) with R2 > 0.8 from genetic loci associated with disease are provided below.*,**,*** indicates candidate genes derived from the same genetic locus.

Gene	Missense Variants	R^2 with RA	Functional Effects
PADI4	Gly55Ser, Val82Ala, Gly112Ala	0.95	Affects PADI4 mRNA stability
PTPN22	Arg620Trp	1	Affects BCR and TCR signalling
IL6R	Asp358Ala	1	Impairs classical IL6R signalling
NCK1	Ala116Val	0.92	-
NFKBIE*	Val194Ala, Pro175Leu	1	Decreased NF-κB activity
TCTE1*	Arg95His	0.94	-
AARS2*	Val730Met	0.88	-
TNFAIP3	Phe127Cys	1	Affects TNFAIP3 mRNA and NFKB activity
WDFY4	Arg1816Gln	0.84	-
RTKN2	Ala288Thr	0.88	Increased mRNA levels and increased NF-κB activity
CD5	Ala471Val	0,9	Increased T-cell proliferation and cytokine release
SH2B3	Trp262Arg	0.86	-
PRKCH**	Val374Ile	1	-
AHNAK2*	Gly1901Ser	1	-
ZPBP2*	Ser151Ile	0.99	-
GSDMB**	Pro298Ser, Gly291Arg	0.99	-
TYK2	Pro110Ala	0.87	-
ICOSLG	Trp353Arg	0.94	-
IRAK1	Phe196Ser, Ser453Leu	0.96	-

Other examples of functional studies to elucidate the mechanism of action of the putative causal variant include PADI4 [49], TNFAIP3 [53], IL6R [54], NFKBIE [55],

CD5 [56] (Table 1) as well as TYK2, CCR6, IL2RA and CD40. The CD40 variant leads to increased cell surface expression of CD40 protein on B cells, leading to enhanced NF-κB pathway activation [57]. Chemokine receptor 6 (CCR6) is considered as a surface marker for Th17 cells [58]. Expression of CCR6 correlates with the polymorphism rs3093024 that was found associated with RA [59]. The CCR6 genotype was also correlated with induced IL-17 levels in the sera of RA patients [59]. Interestingly, an increased number of CCR6+T-cells were identified in peripheral blood, synovial fluid and inflamed synovial tissues of RA patients, highlighting an important role in RA pathogenesis [60]. TNFAIP3 encodes a ubiquitin-modifying enzyme (known as A20) that was identified as a component of the NF-κB signalling pathway [61] (Fig. 2). Three independent polymorphisms were identified in this locus to associate with RA susceptibility [37,62]. One of these variants reduced avidity for transcription factor binding by NF-κB resulting in decreased mRNA expression and reduced A20 protein levels [53]. Moreover, mice lacking A20 in myeloid cells developed spontaneous polyarthritis sharing many features with rheumatoid arthritis [63]. These data indicate the importance of the NF-κB signalling cascade in rheumatoid arthritis and highlight how genes like TNFAIP3 can disturb such immune homeostasis. Another functional variant was identified near IL2RA which encodes the IL2 receptor subunit α. This variant (rs12722495) was shown to correlate with both IL2RA mRNA and protein levels in stimulated monocytes, CD4+ Naïve T cells and memory T-cells [64]. The few attempts at ad-hoc functional characterization of causal variants and causal candidate genes in loci identified in genetic studies of complex diseases has proven to be laborious and challenging, yielding limited advance on our understanding of disease pathogenesis but have highlighted important challenges (i) How do we identify the causal variants (ii) How do we identify the causal genes (iii) which relevant cell types are affected.

5. Pathways involved in RA identified by genetic studies

In anticipation of elucidating this three-pronged puzzle, attempts to identify whether candidate genes are enriched in certain molecular pathways have provided some relevant insight into the mechanisms that are at play. Among the 100 loci associated with RA to date, 377 candidate genes have been identified. Candidate genes are often annotated based on their immune function and their closeness to the lead SNP. More sophisticated ways have been developed and

continue to be implemented [65,66]. Pathway analysis of the 377 candidate genes or 100 prioritised genes using StringDB [67] reveals largely similar pathways as previously described with the top three pathways enriched in RA-associated genes being the JAK-STAT signalling, NF-κB signalling and T-cell receptor signalling pathways (Table 2, Fig. 2).

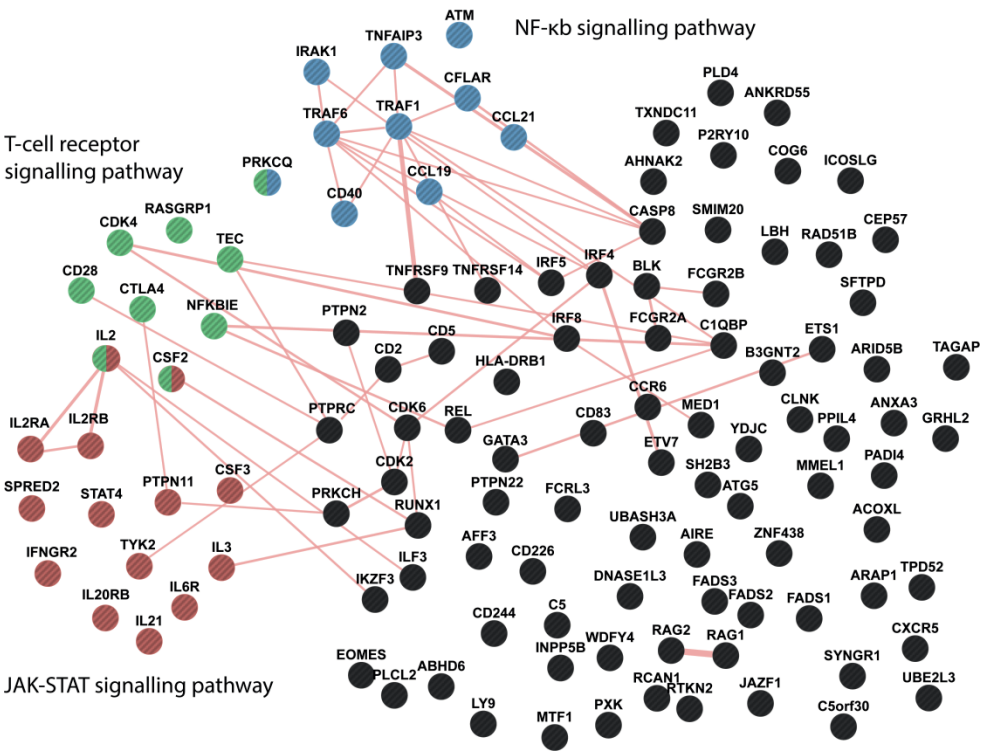

Fig. 2. Interactive overview of top three pathways enriched for rheumatoid arthritis susceptibility genes. Each node represents a susceptibility gene and coloured nodes represent genes in the enriched pathways: JAK-STAT signalling pathway (red), T-cell receptor signalling pathway (green), NF-κb signalling pathway (blue). Protein–protein interaction between susceptibility genes is indicated by red lines. Pathway enrichment analysis was performed in stringDB using KEGG-pathways [117]. GeneMENIA was used for visualisation and protein–protein interactions [118].

Interestingly, several JAK inhibiting drugs are currently under development, with one drug, tofacitinib being approved for treatment of RA [68–71]. In addition, enrichment for B cell and cytokine signalling have also been reported [9]. It is important to map whether “causal” genes or pathways are down or up-regulated

by risk alleles to gain insight into how disease mechanisms operate. However, pathway analyses are limited by what is already known and are often biased towards mechanisms that are mostly studied in the context of common diseases and may therefore limit the possibility of novel hypothesis generating exercises. Bearing this in mind, in order to further our understanding of genes identified and how they play a role in disease, there is a need to move beyond what is already known.

Table 2. Top three pathways identified from candidate genes across 100 loci associated with rheumatoid arthritis (StringDB, Kegg pathway enrichment).

GO ID	Kegg Pathway	# of genes	FDR P	GENES
4630	Jak-STAT signaling pathway	14	2.74E-11	CSF2, CSF3, IL21, IL3, IL20RB, IL2RB, IL6R, IL2, SPRED2, IFNGR2, PTPN11, IL2RA, TYK2, STAT4
4064	NF-kappa B signaling pathway	10	4.75E-9	CCL19, CCL21, PRKCQ, CD40, TNFAIP3, IRAK1, TRAF6, CFLAR, TRAF1, ATM
4660	T cell receptor signaling pathway	9	2.43E-7	IL2, CD28, TEC, CTLA4, CSF2, RASGRP1, CDK4, PRKCQ, CSF2, NFKBIE

6. Non-coding variation, super-enhancers and non-coding RNAs

While coding variants span ~1% of the human genome and explain <10% of the heritability across immune diseases [72], the majority of likely "causal" variants lie in non-coding regions of the genome outside known protein-coding genes and are likely to affect expression of candidate genes [73,74]. In fact, RA-risk SNPs have been found in 44 cis-acting expression quantitative trait loci (cis-eQTL) identified in peripheral blood mononuclear cells [75,76]. Similar observations have been reported for other autoimmune diseases and other cell types, indicating that the quantitative differences of RNA expression with respect to risk alleles may provide clues to disease pathogenesis. Recent studies in CD4+T cells under stimulation conditions revealed cis-eQTLs at 46 genes, 11 of which were previously undetected in peripheral blood mononuclear cells, highlighting the value of well powered cell-type specific analyses to gain novel insights into disease mechanism [77]. There is a growing body of eQTL studies being performed in individual (primary) cell types under basal as well as stimulation conditions [77–89]. Growing evidence suggests that these GWAS signals are enriched in cell-type-specific [90,91], large active regulatory regions of the genome [92–94], known as super-enhancers [95,96]. Based on the analysis of 21

autoimmune GWAS, a recent paper by Farh and colleagues, describes the development of a unique resource for assigning a probability of single nucleotide polymorphisms (SNPs) being causal in disease [97]. 60% of these likely causal variants are located within stimulus and cell-type specific enhancers, identified through both histone modifications and the transcription of noncoding RNAs.

Histone modifications are markers of different chromatin states with methylation or acetylation of specific histones strongly correlating with promoter or enhancer positions and activity [98]. Using this method, Farh *et al*. calculate that the lead SNP is less likely to be the causal variant and is a median 14 kb distance away from the predicted causal SNP. Interestingly, despite the close colocalisation of causal variants to transcription binding motifs, the authors suggest that altering the motif itself to affect binding is unlikely, implying that other as yet unknown mechanisms of mediating their effects remain to be identified.

A more recent study led by Vahedi *et al*. confirms the enrichment of super-enhancers in CD4+T cells in addition to CD56 + NK cells and CD14 + monocytes. CD4+T cells have been repeatedly identified as critical cell types in RA [78,79]. CD4+T cells after being activated and differentiated into distinct effector subtypes play a major role in mediating immune response through the secretion of specific cytokines.

The CD4+T cells carry out multiple functions, ranging from activation of the cells of the innate immune system, B-lymphocytes, cytotoxic T cells, as well as non-immune cells, and also play critical role in the suppression of immune reaction. Importantly, the levels of super-enhancers at GWAS loci detected in CD4+T cells seem to be preferentially affected by the JAK-STAT inhibitor tofacitinib as compared to super-enhancers in non-CD4+T cells, providing indications that genes involved in the disease pathway are likely being targeted [95] and provides hope for such approaches to at least yield valuable drug targets. In addition, these studies help to highlight that genome regulation is dynamic, cell specific and much more complex than previously envisaged [99]. Similar endeavours need to be undertaken for other relevant immune cell (sub)types. Taking into account this complexity, no matter how challenging, will undoubtedly lead us to novel insight as an ever increasing amount of data begins to emerge.

7. Noncoding RNAs as novel candidate genes for rheumatoid arthritis

In recent years, non-coding RNAs have gained much interest as their prevalence in the human genome is much larger than previously anticipated. The ~20,000 protein coding genes occupy less than 2% of total human genome sequence [100]. Not surprisingly, at least 90% of the genome is actively transcribed into noncoding RNAs (ncRNAs), which have no protein coding potentiality [101]. A heterogeneous, novel class of long noncoding RNAs (lncRNAs) with length longer than 200 nucleotides is generally characterized as non-protein transcript [102]. Over the past decade more than 18,000 transcripts have been discovered/annotated as lncRNAs in mammalian transcriptomes [103–105]. Numerous studies have revealed that lncRNAs are believed to form a major proportion of novel transcripts and are known to be involved in number of functionally distinct biological and physiological processes including chromatin remodelling, gene transcription, RNA splicing [106,107] and directly linked to human diseases including various cancers and autoimmune diseases [108–110]. Furthermore, lncRNAs act as a key regulator of inflammatory gene expression by a collaboration involving signal-dependent activation of transcription factors, transcriptional coregulators, and chromatin-modifying factors [111].

Ding and colleagues identified 12 lncRNAs in GWAS regions (LD region defined as R2 > 0.8 with lead SNP) [112]. These lncRNAs were expressed in RNAseq data from the Illumina Human Body Map which consist of 16 human tissue types, including adrenal, adipose, brain, breast, colon, heart, kidney, liver, lung, lymph, ovary, prostate, skeletal muscle, testes, thyroid, and white blood cells. 9 of these were sufficiently far away from protein-coding genes to suggest that they are the putative causal genes in the associated regions. In addition, there are a large number of lncRNAs across the majority of loci identified so far implicating in part that they may be as yet uninvestigated candidate genes. Noncoding RNAs do not easily come up in the list of candidate genes as mostly their functions are not as yet well characterized. Elucidating the roles of lncRNAs and the impact of RA-associated variants have on their function will be an important area of research aimed at elucidating mechanisms of disease susceptibility. Recently, observations of differential regulation of lncRNA pathways relevant to RA have been observed. Various studies show either up or down regulation of lncRNAs after specific immune stimuli [111,113,114]. Specifically, a study evaluated lncRNA expression in CD14 + monocytes from RA patients before and after anti-TNFa or anti-IL6

treatment [115]. 55 lncRNAs were differentially expressed upon TNFa inhibition, while 25 distinct non-overlapping lncRNAs were differentially regulated upon anti-IL6 treatment. Another study performed by Kumar and colleagues has also provided evidence of lncRNA eQTLs from GWAS SNPs (lncRNA eQTLs) further emphasizing the potential role of lncRNAs in the aetiology of disease [116]. Much like enhancers, these lncRNA eQTLs were tissue specific. The average expression of lncRNAs under basal conditions is lower than protein-coding genes. Many more lncRNAs are likely to exist but are as yet undiscovered due to limitations in detection thresholds using quantitative PCR and current depth of RNA sequencing. Further efforts need to be made to further characterize the expression and the function of these lncRNAs in specific cell subsets. There is therefore supporting evidence for a role of lncRNAs in RA and in autoimmunity in general and future studies focussing on this field in autoimmunity is likely going to reveal much about pathogenic mechanisms in disease. In particular, once the role of lncRNAs are better mapped, their consideration as "causal" candidate genes can be included to discover and understand the contribution of genetic findings to disease pathways.

8. Summary and perspectives

In summary, we are currently observing enormous changes in the landscape of moving from genetics to understanding immune function. We are generating new insights in the role of HLA in disease onset through complex laborious functional experiments. In addition, specific cell types including CD4+T cells, B cells as well as novel pathways like the JAK-STAT pathway has been definitively established. Genetic studies have also revealed a striking diversity of molecular pathways to disease, including unexpectedly important contributions of non-coding genetic variation in modulating regulatory elements and immune genes. In particular, advances in technology is increasingly making it possible to (i) prioritise variants more accurately, (ii) prioritise genes more accurately and (iii) prioritise cell-types more accurately. The challenge ahead is to carve out suitable strategies to gain insight into cell-type specific molecular processes and pathways underlying the discovered GWAS signals.

Technologies like mass cytometry, (single cell) gene expression profiling by RNA sequencing and multiplexed functional assays can be leveraged and will enable

the analysis of immune cell function with unprecedented detail and promise not only a deeper understanding of pathogenesis, but also the discovery of novel biomarkers. The large and complex data sets generated by these technologies require specialized approaches for analysis and visualization of results which is a rapidly moving field.

Despite the obvious challenges we have faced and those remaining ahead, it is imperative that we remember that it was barely a decade ago that the first genomes were published and that we are now starting to catalogue a comprehensive list of genetic variation that associates with disease. In contrast to what we mostly expected, that large and obvious changes would be detected in the coding region of the human genome, we are gaining more insight into how our genome is non-linear and dynamic and that taking snapshots of functionality at a given time and under one given condition may have restricted our discovery efforts until now.

Supplementary information

Supplementary Table 1 is available online at the journal of autoimmunity (http://dx.doi.org/10.1016/j.jaut.2015.07.007).

Acknowledgements

This work was supported by the Dutch Arthritis Foundation, a Leiden University Medical Center (LUMC) fellowship (11-1-406), the IMI JU (115142) funded project BeTheCure and European Union (Health-F2-2008-223404) (Seventh Framework Programme integrated project Masterswitch).

References

[1] D.G. Hoy, E. Smith, M. Cross, L. Sanchez-Riera, F.M. Blyth, R. Buchbinder, *et al*. Reflecting on the global burden of musculoskeletal conditions: lessons learnt from the global burden of disease 2010 study and the next steps forward. Ann. Rheum. Dis., 74 (1) (2015 Jan), pp. 4-7

[2] I. Rudan, S. Sidhu, A. Papana, S.J. Meng, Y. Xin-Wei, W. Wang, *et al*. Prevalence of rheumatoid arthritis in low- and middle-income countries: A systematic review and analysis. J. Glob. Health, 5 (1) (2015 Jun), p. 010409

[3] I.B. McInnes, G. Schett The pathogenesis of rheumatoid arthritis. N. Engl. J. Med., 365 (23) (2011 Dec 8), pp. 2205-2219

[4] A. Willemze, L.A. Trouw, R.E. Toes, T.W. Huizinga The influence of ACPA status and characteristics on the course of RA. Nat. Rev. Rheumatol., 8 (3) (2012 Mar), pp. 144-152

[5] K.P. Machold, T.A. Stamm, V.P. Nell, S. Pflugbeil, D. Aletaha, G. Steiner, *et al*. Very recent onset rheumatoid arthritis: clinical and serological patient characteristics associated with radiographic progression over the first years of disease. Rheumatol. Oxf., 46 (2) (2007 Feb), pp. 342-349

[6] A.J. MacGregor, H. Snieder, A.S. Rigby, M. Koskenvuo, J. Kaprio, K. Aho, *et al*. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum., 43 (1) (2000 Jan), pp. 30-37

[7] A.J. Silman, A.J. MacGregor, W. Thomson, S. Holligan, D. Carthy, A. Farhan, *et al*. Twin concordance rates for rheumatoid arthritis: results from a nationwide study. Br. J. Rheumatol., 32 (10) (1993 Oct), pp. 903-907

[8] D. van der Woude, J.J. Houwing-Duistermaat, R.E. Toes, T.W. Huizinga, W. Thomson, J. Worthington, *et al*. Quantitative heritability of anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. Arthritis Rheum., 60 (4) (2009 Apr), pp. 916-923

[9] Y. Okada, D. Wu, G. Trynka, T. Raj, C. Terao, K. Ikari, *et al*. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature, 506 (7488) (2014 Feb 20), pp. 376-381

[10] P. Stastny Association of the B-cell alloantigen DRw4 with rheumatoid arthritis. N. Engl. J. Med., 298 (16) (1978 Apr 20), pp. 869-871

[11] P.K. Gregersen, J. Silver, R.J. Winchester The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum., 30 (11) (1987 Nov), pp. 1205-1213

[12] S. Raychaudhuri, C. Sandor, E.A. Stahl, J. Freudenberg, H.S. Lee, X. Jia, *et al*. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat. Genet., 44 (3) (2012 Mar), pp. 291-296

[13] R.J. Reynolds, A.F. Ahmed, M.I. Danila, L.B. Hughes, P.K. Gregersen, S. Raychaudhuri, *et al*. HLA-DRB1-associated rheumatoid arthritis risk at multiple levels in African Americans: hierarchical lassification systems, amino acid positions, and residues. Arthritis Rheumatol., 66 (12) (2014 Dec), pp. 3274-3282

[14] B. Han, D. Diogo, S. Eyre, H. Kallberg, A. Zhernakova, J. Bowes, *et al*. Fine mapping seronegative and seropositive rheumatoid arthritis to shared and distinct HLA alleles by adjusting for the effects of heterogeneity. Am. J. Hum. Genet., 94 (4) (2014 Apr 3), pp. 522-532

[15] P. Irigoyen, A.T. Lee, M.H. Wener, W. Li, M. Kern, F. Batliwalla, *et al*. Regulation of anti-cyclic citrullinated peptide antibodies in rheumatoid arthritis: contrasting effects of HLA-DR3 and the shared epitope alleles. Arthritis Rheum., 52 (12) (2005 Dec), pp. 3813-3818

[16] K.N. Verpoort, F.A. van Gaalen, A.H. van der Helm-van Mil, G.M. Schreuder, F.C. Breedveld, T.W. Huizinga, *et al*. Association of HLA-DR3 with anti-cyclic citrullinated peptide antibody-negative rheumatoid arthritis. Arthritis Rheum., 52 (10) (2005 Oct), pp. 3058-3062

[17] S. Oka, H. Furukawa, A. Kawasaki, K. Shimada, S. Sugii, A. Hashimoto, *et al*. Protective effect of the HLA-DRB1*13:02 allele in Japanese rheumatoid arthritis patients. PLoS One, 9 (6) (2014), p. e99453

[18] D. van der Woude, B.A. Lie, E. Lundstrom, A. Balsa, A.L. Feitsma, J.J. Houwing-Duistermaat, *et al*. Protection against anti-citrullinated protein antibody-positive rheumatoid arthritis is predominantly associated with HLA-DRB1*1301: a meta-analysis of HLA-DRB1 associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in four European populations. Arthritis Rheum., 62 (5) (2010 May), pp. 1236-1245

[19] H.J. van, D.T. Jansen, S. Polydorides, A.K. Moustakas, M. Bax, A.L. Feitsma, *et al*. Crossreactivity to vinculin and microbes provides a molecular basis for HLA-based protection against rheumatoid arthritis. Nat. Commun., 6 (2015), p. 6681

[20] A.L. Feitsma, J. Worthington, A.H. van der Helm-van Mil, D. Plant, W. Thomson, J. Ursum, *et al*. Protective effect of noninherited maternal HLA-DR antigens on rheumatoid arthritis development. Proc. Natl. Acad. Sci. U. S. A., 104 (50) (2007 Dec 11), pp. 19966-19970

[21] Finishing the euchromatic sequence of the human genome. Nature, 431 (7011) (2004 Oct 21), pp. 931-945

[22] A haplotype map of the human genome. Nature, 437 (7063) (2005 Oct 27), pp. 1299-1320

[23] The International HapMap Project. Nature, 426 (6968) (2003 Dec 18), pp. 789-796

[24] G.R. Abecasis, A. Auton, L.D. Brooks, M.A. DePristo, R.M. Durbin, R.E. Handsaker, *et al*.

An integrated map of genetic variation from 1,092 human genomes. Nature, 491 (7422) (2012 Nov 1), pp. 56-65

[25] B. Howie, J. Marchini, M. Stephens. Genotype imputation with thousands of genomes. G3 (Bethesda), 1 (6) (2011 Nov), pp. 457-470

[26] S. Eyre, J. Bowes, D. Diogo, A. Lee, A. Barton, P. Martin, *et al*. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat. Genet., 44 (12) (2012 Dec), pp. 1336-1340

[27] K.P. Liao, F. Kurreeman, G. Li, G. Duclos, S. Murphy, R. Guzman, *et al*. Associations of autoantibodies, autoimmune risk alleles, and clinical diagnoses from the electronic medical records in rheumatoid arthritis cases and non-rheumatoid arthritis controls. Arthritis Rheum., 65 (3) (2013 Mar), pp. 571-581

[28] Y. Okada, C. Terao, K. Ikari, Y. Kochi, K. Ohmura, A. Suzuki, *et al*. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nat. Genet., 44 (5) (2012 May), pp. 511-516

[29] R. Chen, E.A. Stahl, F.A. Kurreeman, P.K. Gregersen, K.A. Siminovitch, J. Worthington, *et al*. Fine mapping the TAGAP risk locus in rheumatoid arthritis. Genes Immun., 12 (4) (2011 Jun), pp. 314-318

[30] A. Zhernakova, E.A. Stahl, G. Trynka, S. Raychaudhuri, E.A. Festen, L. Franke, *et al*. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS Genet., 7 (2) (2011 Feb), p. e1002004

[31] F. Kurreeman, K. Liao, L. Chibnik, B. Hickey, E. Stahl, V. Gainer, *et al*. Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records. Am. J. Hum. Genet., 88 (1) (2011 Jan 7), pp. 57-69

[32] E.A. Stahl, S. Raychaudhuri, E.F. Remmers, G. Xie, S. Eyre, B.P. Thomson, *et al*. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat. Genet., 42 (6) (2010 Jun), pp. 508-514

[33] S. Raychaudhuri, B.P. Thomson, E.F. Remmers, S. Eyre, A. Hinks, C. Guiducci, *et al*. Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. Nat. Genet., 41 (12) (2009 Dec), pp. 1313-1318

[34] F.A. Kurreeman, N.A. Daha, M. Chang, J.J. Catanese, A.B. Begovich, T.W. Huizinga, *et al*. Association of IL2RA and IL2RB with rheumatoid arthritis: a replication study in a Dutch population. Ann. Rheum. Dis., 68 (11) (2009 Nov), pp. 1789-1790

[35] P.K. Gregersen, C.I. Amos, A.T. Lee, Y. Lu, E.F. Remmers, D.L. Kastner, *et al*. REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. Nat. Genet., 41 (7) (2009 Jul), pp. 820-823

[36] S. Raychaudhuri, E.F. Remmers, A.T. Lee, R. Hackett, C. Guiducci, N.P. Burtt, *et al*. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. Nat. Genet., 40 (10) (2008 Oct), pp. 1216-1223

[37] R.M. Plenge, C. Cotsapas, L. Davies, A.L. Price, P.I. de Bakker, J. Maller, *et al*. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat. Genet., 39 (12) (2007 Dec), pp. 1477-1482

[38] F.A. Kurreeman, L. Padyukov, R.B. Marques, S.J. Schrodi, M. Seddighzadeh, G. Stoeken-Rijsbergen, *et al*. A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. PLoS Med., 4 (9) (2007 Sep), p. e278

[39] E.F. Remmers, R.M. Plenge, A.T. Lee, R.R. Graham, G. Hom, T.W. Behrens, *et al*. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N. Engl. J. Med., 357 (10) (2007 Sep 6), pp. 977-986

[40] R.M. Plenge, M. Seielstad, L. Padyukov, A.T. Lee, E.F. Remmers, B. Ding, *et al*. TRAF1-C5 as a risk locus for rheumatoid arthritis–a genomewide study. N. Engl. J. Med., 357 (12) (2007 Sep 20), pp. 1199-1209

[41] E.W. Karlson, L.B. Chibnik, J. Cui, R.M. Plenge, R.J. Glass, N.E. Maher, *et al*. Associations between human leukocyte antigen, PTPN22, CTLA4 genotypes and rheumatoid arthritis phenotypes of

autoantibody status, age at diagnosis and erosions in a large cohort study. Ann. Rheum. Dis., 67 (3) (2008 Mar), pp. 358-363

[42] R.M. Plenge, L. Padyukov, E.F. Remmers, S. Purcell, A.T. Lee, E.W. Karlson, *et al*. Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. Am. J. Hum. Genet., 77 (6) (2005 Dec), pp. 1044-1060

[43] E.A. Stahl, D. Wegmann, G. Trynka, J. Gutierrez-Achury, R. Do, B.F. Voight, *et al*. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nat. Genet., 44 (5) (2012 May), pp. 483-489

[44] J.F. Cloutier, A. Veillette. Cooperative inhibition of T-cell antigen receptor signaling by a complex between a kinase and a phosphatase. J. Exp. Med., 189 (1) (1999 Jan 4), pp. 111-121

[45] M. Rieck, A. Arechiga, S. Onengut-Gumuscu, C. Greenbaum, P. Concannon, J.H. Buckner. Genetic variation in PTPN22 corresponds to altered function of T and B lymphocytes. J. Immunol., 179 (7) (2007 Oct 1), pp. 4704-4710

[46] D.J. Rawlings, X. Dai, J.H. Buckner. The role of PTPN22 risk variant in the development of autoimmunity: finding common ground between mouse and human. J. Immunol., 194 (7) (2015 Apr 1), pp. 2977-2984

[47] G.L. Burn, L. Svensson, C. Sanchez-Blanco, M. Saini, A.P. Cope. Why is PTPN22 a good candidate susceptibility gene for autoimmune disease? FEBS Lett., 585 (23) (2011 Dec 1), pp. 3689-3698

[48] J. Zhang, N. Zahir, Q. Jiang, H. Miliotis, S. Heyraud, X. Meng, *et al*. The autoimmune disease-associated PTPN22 variant promotes calpain-mediated Lyp/Pep degradation associated with lymphocyte and dendritic cell hyperresponsiveness. Nat. Genet., 43 (9) (2011 Sep), pp. 902-907

[49] A. Suzuki, R. Yamada, X. Chang, S. Tokuhiro, T. Sawada, M. Suzuki, *et al*. Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. Nat. Genet., 34 (4) (2003 Aug), pp. 395-402

[50] T.W. Huizinga, C.I. Amos, A.H. van der Helm-van Mil, W. Chen, F.A. van Gaalen, D. Jawaheer, *et al*. Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins. Arthritis Rheum., 52 (11) (2005 Nov), pp. 3433-3438

[51] V. Brinkmann, U. Reichard, C. Goosmann, B. Fauler, Y. Uhlemann, D.S. Weiss, *et al*. Neutrophil extracellular traps kill bacteria. Science, 303 (5663) (2004 Mar 5), pp. 1532-1535

[52] H.H. Chang, N. Dwivedi, A.P. Nicholas, I.C. Ho. The W620 polymorphism of PTPN22 disrupts its interaction with PAD4 and enhances citrullination and NETosis. Arthritis Rheumatol. (2015 May 27), 10.1002/art.39215

[53] I. Adrianto, F. Wen, A. Templeton, G. Wiley, J.B. King, C.J. Lessard, *et al*. Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. Nat. Genet., 43 (3) (2011 Mar), pp. 253-258

[54] R.C. Ferreira, D.F. Freitag, A.J. Cutler, J.M. Howson, D.B. Rainbow, D.J. Smyth, *et al*. Functional IL6R 358Ala allele impairs classical IL-6 receptor signaling and influences risk of diverse inflammatory diseases. PLoS Genet., 9 (4) (2013 Apr), p. e1003444

[55] K. Myouzen, Y. Kochi, Y. Okada, C. Terao, A. Suzuki, K. Ikari, *et al*. Functional variants in NFKBIE and RTKN2 involved in activation of the NF-kappaB pathway are associated with rheumatoid arthritis in Japanese. PLoS Genet., 8 (9) (2012 Sep), p. e1002949

[56] M.C. Cenit, M. Martinez-Florensa, M. Consuegra, L. Bonet, E. Carnero-Montoro, N. Armiger, *et al*. Analysis of ancestral and functionally relevant CD5 variants in systemic lupus erythematosus patients. PLoS One, 9 (11) (2014), p. e113090

[57] G. Li, D. Diogo, D. Wu, J. Spoonamore, V. Dancik, L. Franke, *et al*. Human genetics in rheumatoid arthritis guides a high-throughput drug screen of the CD40 signaling pathway. PLoS Genet., 9 (5) (2013 May), p. e1003487

[58] J.D. Haas, F.H. Gonzalez, S. Schmitz, V. Chennupati, L. Fohse, E. Kremmer, *et al*. CCR6 and NK1.1 distinguish between IL-17A and IFN-gamma-producing gammadelta effector T cells. Eur. J. Immunol., 39 (12) (2009 Dec), pp. 3488-3497

[59] Y. Kochi, Y. Okada, A. Suzuki, K. Ikari, C. Terao, A. Takahashi, *et al*. A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. Nat. Genet., 42 (6) (2010 Jun), pp. 515-519

[60] J.P. van Hamburg, P.S. Asmawidjaja, N. Davelaar, A.M. Mus, E.M. Colin, J.M. Hazes, *et al*. Th17 cells, but not Th1 cells, from patients with early rheumatoid arthritis are potent inducers of matrix metalloproteinases and proinflammatory cytokines upon synovial fibroblast interaction, including autocrine interleukin-17A production. Arthritis Rheum., 63 (1) (2011 Jan), pp. 73-83

[61] H.Y. Song, M. Rothe, D.V. Goeddel. The tumor necrosis factor-inducible zinc finger protein A20 interacts with TRAF1/TRAF2 and inhibits NF-kappaB activation. Proc. Natl. Acad. Sci. U. S. A., 93 (13) (1996 Jun 25), pp. 6721-6725

[62] G. Orozco, A. Hinks, S. Eyre, X. Ke, L.J. Gibbons, J. Bowes, *et al*. Combined effects of three independent SNPs greatly increase the risk estimate for RA at 6q23. Hum. Mol. Genet., 18 (14) (2009 Jul 15), pp. 2693-2699

[63] M. Matmati, P. Jacques, J. Maelfait, E. Verheugen, M. Kool, M. Sze, *et al*. A20 (TNFAIP3) deficiency in myeloid cells triggers erosive polyarthritis resembling rheumatoid arthritis. Nat. Genet., 43 (9) (2011 Sep), pp. 908-912

[64] C.A. Dendrou, V. Plagnol, E. Fung, J.H. Yang, K. Downes, J.D. Cooper, *et al*. Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable human bioresource. Nat. Genet., 41 (9) (2009 Sep), pp. 1011-1015

[65] S. Raychaudhuri, R.M. Plenge, E.J. Rossin, A.C. Ng, S.M. Purcell, P. Sklar, *et al*. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet., 5 (6) (2009 Jun), p. e1000534

[66] T.H. Pers, J.M. Karjalainen, Y. Chan, H.J. Westra, A.R. Wood, J. Yang, *et al*. Biological interpretation of genome-wide association studies using predicted gene functions. Nat. Commun., 6 (2015), p. 5890

[67] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, *et al*. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res., 43 (Database issue) (2015 Jan), pp. D447-D452

[68] K. Garber. Pfizer's JAK inhibitor sails through phase 3 in rheumatoid arthritis. Nat. Biotechnol., 29 (6) (2011 Jun), pp. 467-468

[69] K. Nishimura, J. Saegusa, F. Matsuki, K. Akashi, G. Kageyama, A. Morinobu. Tofacitinib facilitates the expansion of myeloid-derived suppressor cells and ameliorates arthritis in SKG mice. Arthritis Rheumatol., 67 (4) (2015 Apr), pp. 893-902

[70] D.L. Boyle, K. Soma, J. Hodge, A. Kavanaugh, D. Mandel, P. Mease, *et al*. The JAK inhibitor tofacitinib suppresses synovial JAK1-STAT signalling in rheumatoid arthritis. Ann. Rheum. Dis., 74 (6) (2015 Jun), pp. 1311-1316

[71] K. Garber. Pfizer's first-in-class JAK inhibitor pricey for rheumatoid arthritis market. Nat. Biotechnol., 31 (1) (2013 Jan), pp. 3-4

[72] A. Gusev, S.H. Lee, G. Trynka, H. Finucane, B.J. Vilhjalmsson, H. Xu, *et al*. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am. J. Hum. Genet., 95 (5) (2014 Nov 6), pp. 535-552

[73] D.L. Nicolae, E. Gamazon, W. Zhang, S. Duan, M.E. Dolan, N.J. Cox. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet., 6 (4) (2010 Apr), p. e1000888

[74] A.C. Nica, S.B. Montgomery, A.S. Dimas, B.E. Stranger, C. Beazley, I. Barroso, *et al*. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet., 6 (4) (2010 Apr), p. e1000895

[75] H.J. Westra, M.J. Peters, T. Esko, H. Yaghootkar, C. Schurmann, J. Kettunen, *et al*. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet., 45 (10) (2013 Oct), pp. 1238-1243

[76] P.C. Dubois, G. Trynka, L. Franke, K.A. Hunt, J. Romanos, A. Curtotti, *et al*. Multiple common variants for celiac disease influencing immune gene expression. Nat. Genet., 42 (4) (2010 Apr), pp. 295-302

[77] X. Hu, H. Kim, T. Raj, P.J. Brennan, G. Trynka, N. Teslovich, *et al*. Regulation of gene expression in autoimmune disease loci and the genetic basis of proliferation in CD4+ effector memory T cells. PLoS Genet., 10 (6) (2014 Jun), p. e1004404

[78] X. Hu, H. Kim, E. Stahl, R. Plenge, M. Daly, S. Raychaudhuri. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. Am. J. Hum. Genet., 89 (4) (2011 Oct 7), pp. 496-506

[79] T. Raj, K. Rothamel, S. Mostafavi, C. Ye, M.N. Lee, J.M. Replogle, *et al*. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. Science, 344 (6183) (2014 May 2), pp. 519-523

[80] S. Kim, J. Becker, M. Bechheim, V. Kaiser, M. Noursadeghi, N. Fricker, *et al*. Characterizing the genetic basis of innate immune response in TLR4-activated human monocytes. Nat. Commun., 5 (2014), p. 5236

[81] M. Rotival, T. Zeller, P.S. Wild, S. Maouche, S. Szymczak, A. Schillert, *et al*. Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. PLoS Genet., 7 (12) (2011 Dec), p. e1002367

[82] L. Franke, R.C. Jansen. eQTL analysis in humans. Methods Mol. Biol., 573 (2009), pp. 311-328

[83] H.J. Westra, D. Arends, T. Esko, M.J. Peters, C. Schurmann, K. Schramm, *et al*. Cell Specific eQTL Analysis without Sorting Cells. PLoS Genet., 11 (5) (2015 May), p. e1005223

[84] B.P. Fairfax, J.C. Knight. Genetics of gene expression in immunity to infection. Curr. Opin. Immunol., 30 (2014 Oct), pp. 63-71

[85] B.P. Fairfax, P. Humburg, S. Makino, V. Naranbhai, D. Wong, E. Lau, *et al*. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science, 343 (6175) (2014 Mar 7), p. 1246949

[86] P.L. De Jager, N. Hacohen, D. Mathis, A. Regev, B.E. Stranger, C. Benoist. ImmVar project: Insights and design considerations for future studies of “healthy” immune variation. Semin. Immunol., 27 (1) (2015 Feb), pp. 51-57

[87] M.N. Lee, C. Ye, A.C. Villani, T. Raj, W. Li, T.M. Eisenhaure, *et al*. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. Science, 343 (6175) (2014 Mar 7), p. 1246980

[88] J.K. Pickrell, J.C. Marioni, A.A. Pai, J.F. Degner, B.E. Engelhardt, E. Nkadori, *et al*. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature, 464 (7289) (2010 Apr 1), pp. 768-772

[89] S.B. Montgomery, M. Sammeth, M. Gutierrez-Arcelus, R.P. Lach, C. Ingle, J. Nisbett, *et al*. Transcriptome genetics using second generation sequencing in a Caucasian population. Nature, 464 (7289) (2010 Apr 1), pp. 773-777

[90] G. Trynka, S. Raychaudhuri. Using chromatin marks to interpret and localize genetic associations to complex human traits and diseases. Curr. Opin. Genet. Dev., 23 (6) (2013 Dec), pp. 635-641

[91] G. Trynka, C. Sandor, B. Han, H. Xu, B.E. Stranger, X.S. Liu, *et al*. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat. Genet., 45 (2) (2013 Feb), pp. 124-130

[92] An integrated encyclopedia of DNA elements in the human genome. Nature, 489 (7414) (2012 Sep 6), pp. 57-74

[93] M.T. Maurano, R. Humbert, E. Rynes, R.E. Thurman, E. Haugen, H. Wang, *et al*. Systematic localization of common disease-associated variation in regulatory DNA. Science, 337 (6099) (2012 Sep 7), pp. 1190-1195

[94] R.E. Thurman, E. Rynes, R. Humbert, J. Vierstra, M.T. Maurano, E. Haugen, *et al*. The accessible chromatin landscape of the human genome. Nature, 489 (7414) (2012 Sep 6), pp. 75-82

[95] G. Vahedi, Y. Kanno, Y. Furumoto, K. Jiang, S.C. Parker, M.R. Erdos, *et al*. Super-enhancers delineate disease-associated regulatory nodes in T cells. Nature, 520 (7548) (2015 Apr 23), pp. 558-562

[96] D. Hnisz, B.J. Abraham, T.I. Lee, A. Lau, V. Saint-Andre, A.A. Sigova, *et al*. Super-enhancers in the control of cell identity and disease. Cell, 155 (4) (2013 Nov 7), pp. 934-947

[97] K.K. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W.J. Housley, S. Beik, *et al*. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature, 518 (7539) (2015 Feb 19), pp. 337-343

[98] J. Ernst, P. Kheradpour, T.S. Mikkelsen, N. Shoresh, L.D. Ward, C.B. Epstein, *et al*. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature, 473 (7345) (2011 May 5), pp. 43-49

[99] A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, *et al*. Integrative analysis of 111 reference human epigenomes. Nature, 518 (7539) (2015 Feb 19), pp. 317-330

[100] P. Carninci, T. Kasukawa, S. Katayama, J. Gough, M.C. Frith, N. Maeda, *et al*. The transcriptional landscape of the mammalian genome. Science, 309 (5740) (2005 Sep 2), pp. 1559-1563

[101] E. Birney, J.A. Stamatoyannopoulos, A. Dutta, R. Guigo, T.R. Gingeras, E.H. Margulies, *et al*. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature, 447 (7146) (2007 Jun 14), pp. 799-816

[102] D. Bu, K. Yu, S. Sun, C. Xie, G. Skogerbo, R. Miao, *et al*. NONCODE v3.0: integrative annotation of long noncoding RNAs. Nucleic Acids Res., 40 (Database issue) (2012 Jan), pp. D210-D215

[103] D. Bhartiya, K. Pal, S. Ghosh, S. Kapoor, S. Jalali, B. Panwar, *et al*. lncRNome: a comprehensive knowledgebase of human long noncoding RNAs. Database (Oxford), 2013 (2013), p. bat034

[104] J.E. Wilusz, H. Sunwoo, D.L. Spector. Long noncoding RNAs: functional surprises from the RNA world. Genes Dev., 23 (13) (2009 Jul 1), pp. 1494-1504

[105] J. Zhao, B.K. Sun, J.A. Erwin, J.J. Song, J.T. Lee. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science, 322 (5902) (2008 Oct 31), pp. 750-756

[106] T.R. Mercer, M.E. Dinger, J.S. Mattick. Long non-coding RNAs: insights into functions. Nat. Rev. Genet., 10 (3) (2009 Mar), pp. 155-159

[107] K.C. Wang, H.Y. Chang. Molecular mechanisms of long noncoding RNAs. Mol. Cell, 43 (6) (2011 Sep 16), pp. 904-914

[108] J. Li, Z. Xuan, C. Liu. Long non-coding RNAs and complex human diseases. Int. J. Mol. Sci., 14 (9) (2013), pp. 18790-18808

[109] B. Gupta, R.D. Hawkins. Epigenomics of autoimmune diseases. Immunol. Cell Biol., 93 (3) (2015 Mar), pp. 271-276

[110] T. Gutschner, S. Diederichs. The hallmarks of cancer: a long non-coding RNA point of view. RNA Biol., 9 (6) (2012 Jun), pp. 703-719

[111] S. Carpenter, D. Aiello, M.K. Atianand, E.P. Ricci, P. Gandhi, L.L. Hall, *et al*. A long noncoding RNA mediates both activation and repression of immune response genes. Science, 341 (6147) (2013 Aug 16), pp. 789-792

[112] J. Ding, S. Eyre, J. Worthington. Genetics of RA susceptibility, what comes next? RMD Open. 2015; 1(1): e000028.

[113] H. Cui, N. Xie, Z. Tan, S. Banerjee, V.J. Thannickal, E. Abraham, *et al*. The human long noncoding RNA lnc-IL7R regulates the inflammatory response. Eur. J. Immunol., 44 (7) (2014 Jul), pp. 2085-2095

[114] A.P. Mao, J. Shen, Z. Zuo. Expression and regulation of long noncoding RNAs in TLR4 signaling in mouse macrophages. BMC Genomics, 16 (2015), p. 45

[115] N. Muller, F. Doring, M. Klapper, K. Neumann, D.M. Schulte, K. Turk, *et al*. Interleukin-6 and tumour necrosis factor-alpha differentially regulate lincRNA transcripts in cells of the innate immune system in vivo in human subjects with rheumatoid arthritis. Cytokine, 68 (1) (2014 Jul), pp. 65-68

[116] V. Kumar, H.J. Westra, J. Karjalainen, D.V. Zhernakova, T. Esko, B. Hrdlickova, *et al*. Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. PLoS Genet., 9 (1) (2013), p. e1003201

[117] L.J. Jensen, M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, *et al*. STRING 8–a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res., 37 (Database issue) (2009 Jan), pp. D412-D416

[118] D. Warde-Farley, S.L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, *et al*. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res., 38 (Web Server issue) (2010 Jul), pp. W214-W220

Chapter 4

Comment on "Functional Analysis of a Complement Polymorphism (rs17611) Associated with Rheumatoid Arthritis"

J Immunol. 2015 Jul 1;195(1):3-4. doi: 10.4049/jimmunol.1500822.

Messemaker T, Toes RE, Mikkers HM, Kurreeman F

Comment on "Functional Analysis of a Complement Polymorphism (rs17611) Associated with Rheumatoid Arthritis"

In their published article, Giles *et al*. (1) explored functional consequences of rs17611 in the TRAF1-C5 region and propose a mechanism for its contribution to rheumatoid arthritis (RA) pathology, as this SNP would associate with RA. To substantiate the claim that the SNP associates with RA, the authors referred to the following studies: Refs. 2 and 3.

However, the data presented by Kurreeman *et al*. do not indicate a significant association (p = 0.19) in 544 subjects (Ref. 2, see Table I). In addition, haplotype block analyses of this risk locus shows that block 2 is significantly associated with RA, while block 3 (which includes rs17611) is not (Ref. 2, see Fig. 1). Similar data by Plenge *et al*. (3) across 3372 subjects support this lack of association. Recent genome-wide association studies data across 55,000 European subjects (4) provide accurate estimations of risk across hundreds of SNPs in this locus.

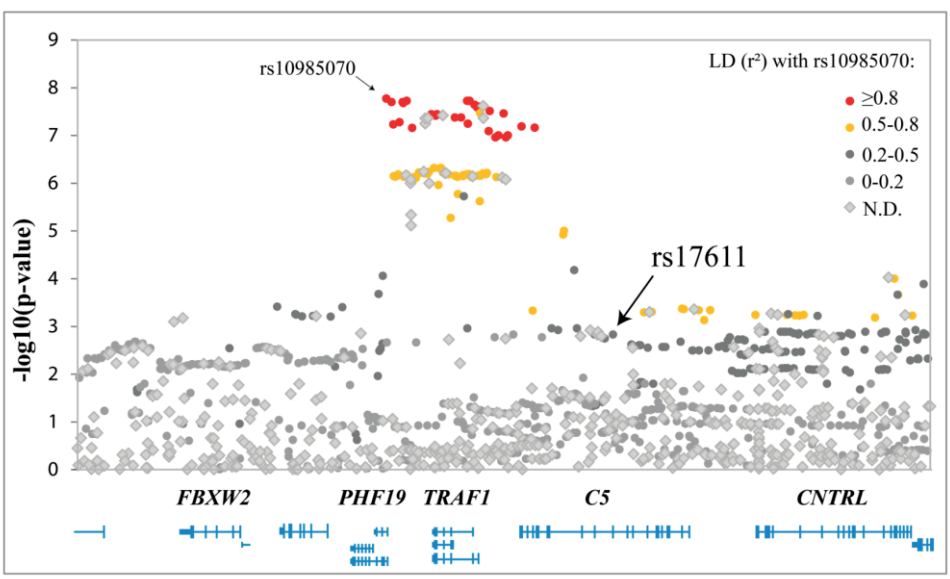

Figure 1. Regional association plot for the TRAF1-C5 region. The *x*-axis shows the chromosomal position of the all queried SNPs located on chromosome 9p33p34 over a region of 500 kb. The *y*-axis displays the –log10 (p value) of associations with RA. The p values were a result of analysis on the European population as part of the Okada *et al*. paper (4): 14,361 RA cases and 43,923 controls from 18 studies of Europeans descent. Pairwise LD values (r^2) were calculated from individuals of the 1000 Genomes Project (CEU) relative to the highest associating SNP rs10985070 using SNAP (6). N.D., LD for SNP is not determined.

Again, no signal of association for rs17611 has been demonstrated in the available published datasets originating from this study (Fig. 1), which also shows that rs17611 is in relatively low linkage disequilibrium (LD) with the lead SNP rs10985070 (r^2 = 0.4). Moreover, independent secondary association signals were not observed in the TRAF1-C5 locus in a study by Eyre *et al*. (5), excluding the possibility that rs17611 could represent a second hit at this locus. On the basis of the published genetic data, we believe caution should be taken in implicating the effects of rs17611 in relation to the immunological mechanism underlying the genetic risk of TRAF1-C5 to RA.

References

1. Giles J. L., E. Choy, C. van den Berg, B. P. Morgan, C. L. Harris. 2015. Functional analysis of a complement polymorphism (rs17611) associated with rheumatoid arthritis. J. Immunol. 194: 3029–3034.
2. Kurreeman F. A. S., L. Padyukov, R. B. Marques, S. J. Schrodi, M. Seddighzadeh, G. Stoeken-Rijsbergen, A. H. M. van der Helm-van Mil, C. F. Allaart, W. Verduyn, J. Houwing-Duistermaat, *et al*. 2007. A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. PLoS Med. 4: e278.
3. Plenge R. M., M. Seielstad, L. Padyukov, A. T. Lee, E. F. Remmers, B. Ding, A. Liew, H. Khalili, A. Chandrasekaran, L. R. Davies, *et al*. 2007. TRAF1-C5 as a risk locus for rheumatoid arthritis—a genomewide study. N. Engl. J. Med. 357: 1199–1209.
4. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., *et al*. 2013. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506: 376–381.
5. Eyre S., J. Bowes, D. Diogo, A. Lee, A. Barton, P. Martin, A. Zhernakova, E. Stahl, S. Viatte, K. McAllister, *et al*, Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate; Wellcome Trust Case Control Consortium. 2012. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat. Genet. 44: 1336–1340.
6. Johnson A. D., R. E. Handsaker, S. L. Pulit, M. M. Nizzari, C. J. O'Donnell, P. I. W. de Bakker. 2008. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24: 2938–2939.

Response to Comment on "Functional Analysis of a Complement Polymorphism (rs17611) Associated with Rheumatoid Arthritis"

Joanna L. Giles, Ernest Choy, Carmen van den Berg, B. Paul Morgan and Claire L. Harris

In their *Comment*, Messemaker *et al*. discuss genetic data that show that block 2 within the TRAF1-C5 locus is a risk for rheumatoid arthritis (RA), while block 3 is

not associated. We acknowledge these data, which illustrate that, when taken in isolation, block 3 does not impact disease risk. It is clear that the genetic data are robust (1); however, there is no question that the rs17611 single nucleotide polymorphism (SNP) in C5 results in an amino acid change in C5 that alters its rate of cleavage by elastase: an enzyme present at high levels in neutrophil-driven or neutrophil-associated diseases such as RA. We believe that it would be important and interesting to further analyze whether this functional polymorphism in block 3, which affects the proinflammatory capacity of C5, has any impact on the risk haplotype in block 2. While genome-wide association studies have identified a number of independent secondary association signals, these are not exclusive (2); the strong functional data that we demonstrate with the C5-V802I variant suggest that potential interactions should be specifically tested in future studies to confirm or exclude them.

In our article (3), we reference other papers where the specific rs17611 SNP has been linked to different diseases. Chai *et al*. (4) show an association ($p < 0.007$) of rs17611 (and its linked set) with periodontal disease; Woehrl *et al*. (5) show an association ($p < 0.002$) of rs17611 with outcome in pneumococcal meningitis; Hoke *et al*. (6) show an association ($p < 0.01$) of rs17611 with adverse cardiovascular outcome; and Greisenegger *et al*. (7) show an association ($p < 0.005$) of rs17611 with risk for ischemic stroke. Together, these linkages provided a strong rationale to explore the functional consequences of the rs17611 SNP on C5 activities, the major focus of our paper.

First, we show that the rs17611 SNP is associated with clear differences in C5 turnover and elevated levels of the proinflammatory product C5a in healthy individuals and RA patients; our finding warrants further investigation of the impact of this polymorphism on risk associated with block 2. We go on to define the mechanism by which the single amino acid change provokes these differences. The genetic data do not detract from the importance of our demonstration that the polymorphism impacts C5 turnover and increases plasma levels of the proinflammatory molecule C5a, and of our description of the atypical route by which this is achieved (3).

References

1. Kurreeman F. A., L. Padyukov, R. B. Marques, S. J. Schrodi, M. Seddighzadeh, G. Stoeken-Rijsbergen, A. H. van der Helm-van Mil, C. F. Allaart, W. Verduyn, J. Houwing-Duistermaat, *et al*. 2007. A

candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. PLoS Med. 4: e278.

2. Eyre S., J. Bowes, D. Diogo, A. Lee, A. Barton, P. Martin, A. Zhernakova, E. Stahl, S. Viatte, K. McAllister, *et al*, Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate; Wellcome Trust Case Control Consortium. 2012. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat. Genet. 44: 1336–1340.
3. Giles J. L., E. Choy, C. van den Berg, B. P. Morgan, C. L. Harris. 2015. Functional analysis of a complement polymorphism (rs17611) associated with rheumatoid arthritis. J. Immunol. 194: 3029–3034.
4. Chai L., Y. Q. Song, K. Y. Zee, W. K. Leung. 2010. Single nucleotide polymorphisms of complement component 5 and periodontitis. J. Periodontal Res. 45: 301–308.
5. Woehrl B., M. C. Brouwer, C. Murr, S. G. Heckenberg, F. Baas, H. W. Pfister, A. H. Zwinderman, B. P. Morgan, S. R. Barnum, A. van der Ende, *et al*. 2011. Complement component 5 contributes to poor disease outcome in humans and mice with pneumococcal meningitis. J. Clin. Invest. 121: 3943–3953.
6. Hoke M., W. Speidl, M. Schillinger, E. Minar, S. Zehetmayer, M. Schönherr, O. Wagner, C. Mannhalter. 2012. Polymorphism of the complement 5 gene and cardiovascular outcome in patients with atherosclerosis. Eur. J. Clin. Invest. 42: 921–926.
7. Greisenegger S., S. Zehetmayer, P. Bauer, G. Endler, J. Ferrari, W. Lang, M. Janisiw, L. Steiner, S. Cheng, W. Lalouschek, C. Mannhalter. 2009. Polymorphisms in inflammatory genes and the risk of ischemic stroke and transient ischemic attack: results of a multilocus genotyping assay. Clin. Chem. 55: 134–138.

Chapter 5

A novel long non-coding RNA in the rheumatoid arthritis risk locus *TRAF1-C5* influences C5 mRNA levels

Genes Immun. 2016 Mar;17(2):85-92. doi: 10.1038/gene.2015.54.

Messemaker TC, Frank-Bertoncelj M, Marques RB, Adriaans A, Bakker AM, Daha N, Gay S, Huizinga TW, Toes RE, Mikkers HM, Kurreeman F

Abstract

Long non-coding RNAs (lncRNAs) can regulate the transcript levels of genes in the same genomic region. These locally acting lncRNAs have been found deregulated in human disease and some have been shown to harbour quantitative trait loci (eQTLs) in autoimmune diseases. However, lncRNAs linked to the transcription of candidate risk genes in loci associated to rheumatoid arthritis (RA) have not yet been identified. The *TRAF1* and *C5* risk locus shows evidence of multiple eQTLs and transcription of intergenic non-coding sequences. Here, we identified a non-coding transcript (C5T1lncRNA) starting in the 3' untranslated region (UTR) of *C5*. RA-relevant cell types express *C5T1lncRNA* and RNA levels are further enhanced by specific immune stimuli. *C5T1lncRNA* is expressed predominantly in the nucleus and its expression correlates positively with C5 mRNA in various tissues (*P*=0.001) and in peripheral blood mononuclear cells (*P*=0.02) indicating transcriptional co-regulation. Knockdown results in a concurrent decrease in C5 mRNA levels but not of other neighbouring genes. Overall, our data show the identification of a novel lncRNA *C5T1lncRNA* that is fully located in the associated region and influences transcript levels of *C5*, a gene previously linked to RA pathogenesis.

Introduction

A large proportion of the mammalian genome is transcribed into non-coding RNAs (ncRNAs), which are thought to be equally important for normal development and physiology as coding genes and have been found deregulated in human disease.[1] ncRNAs with a length of >200 nucleotides are termed long non-coding RNA genes (lncRNAs).[2] The identification of lncRNA genes in the human genome has increased over the years, and was recently estimated at, minimally, 50 000 genes.[3] A variety of mechanisms by which lncRNAs positively or negatively regulate coding genes are known although our understanding is likely going to increase in the coming years.[4,5] These include miRNA sponges,[6] recruitment of proteins that directly enhance or interfere with transcription,[7] and recruitment of chromatin modifiers like polycomb repressor complexes 1 (PRC1)[8] and 2 (PRC2),[9] histone demethylases (LSD1),[10] and methyl transferases (G9a).[11,12]

As certain locally acting lncRNAs have been shown to control transcript levels of neighbouring genes,[13] deregulation of the regulatory lncRNAs may contribute to

disease by affecting the expression of adjacent coding genes. However, in spite of the vast amount of lncRNA genes in the human genome, the percentage of lncRNAs directly linked to disease is extraordinarily low in comparison to coding genes.[1] Genome wide association studies (GWAS) have identified tens of thousands of genomic loci that are associated to complex, multifactorial disorders.[14] One example of a ncRNA disease gene is the lncRNA *ANRIL*. *ANRIL* is located in a 50-kb genetic risk region of atherosclerosis and regulates transcription *in trans* by recruiting PRC2 to genome-wide ALU elements affecting various atherogenic cell functions as cell proliferation, cell adhesion and apoptosis.[15]

We and others have previously identified the *TRAF1-C5* region associated with rheumatoid arthritis (RA), a common disease of autoimmune origin.[16,17] The *TRAF1-C5* region of association is relatively large containing multiple immune-related candidate genes. *TRAF1* and *C5* are excellent, evidence based, candidate RA disease genes. TRAF1 negatively regulates TNFa signalling,[18] which is a successful therapeutic target for RA.[19] Increased levels of C5 have been found in inflamed joints of RA patients,[20] and C5-deficient mice are resistant to the development of collagen-induced arthritis.[21] In addition, both genes have a role in innate immune responses.[22,23] Deciphering disease-associated risk loci to reveal causal genes and mechanisms has remained a challenge with the majority of disease-associated single-nucleotide polymorphisms located in non-coding regions.[24,25] However, recent studies have reported that such regions may contain regulatory RNAs.[26,27] In this study, we performed expression analysis of the intergenic region between *TRAF1* and *C5*. We identified a lncRNA in this region and studied its effect on neighbouring genes in RA relevant cell types.

Results

The intergenic region of the RA risk locus TRAF1 and C5 is transcribed

The number of RA-associated loci have reached over 100, with few causal genes identified.[27,28] Recent insight suggests the importance of regulatory RNAs in GWAS loci.[27] Regulatory lncRNAs can influence the expression of multiple genes in a locus.[11,12] As we previously identified the *TRAF1-C5* locus as a susceptibility locus for RA, we decided to further pursue characterisation of this locus. Analysis of the *TRAF1-C5* risk region in the University of California Santa Cruz (UCSC),

genome browser did not show the presence of non-coding genes and neither did exploration of various lncRNA databases including LNCipedia[29] and LncRBase.[30] As most commonly used RNA-sequencing approaches lack sensitivity,[31,32] we continued expression analysis of the intergenic region of *C5* and *TRAF1* in a panel of 22 tissues using intergenic primer sets (Figure 1a). We observed expression of the intergenic region in all tissues analysed (Figure 1b). Highest expression was seen in liver although in general, intergenic expression was rather low in comparison with coding genes. Before a full characterisation of the transcript, we wished to confirm that the identified transcript of >200 bp is the result of legitimate transcription. As the transcript levels were the highest in liver, we inhibited transcription in a hepatocyte cell line (Huh7) with α-amanitin. Concentrations of α-amanitin (50 µM) that specifically inhibit RNA polymerase II (RNAPII)[33] resulted in a decrease of the intergenic transcript indicating that transcription of the intergenic region is mediated by RNAPII (Figure 1c).

The intergenic region of TRAF1 and C5 contains a lncRNA gene

We next wished to further characterise the nature of the intergenic transcript. Intergenic non-coding genes can be uni- or bidirectionally transcribed. For example, non-coding RNAs transcribed from enhancer elements (eRNA) are most of the time generated by bidirectional transcription.[34,35] Therefore, we first analysed the direction of transcription by strand-specific cDNA synthesis. Here, we could only detect RNA originating from the lagging strand (data not shown), aligning transcriptional directionality in the same orientation as that of the neighbouring genes *TRAF1* and *C5*. Consequently, we reasoned that the identified intergenic transcribed sequence could still be a part of the upstream-located *C5* gene or downstream-located *TRAF1* gene. Identifying the 5′ and 3′ ends of the transcript would allow discrimination between an independent transcript or an alternative TRAF1 or C5 transcript. Like coding genes, RNAPII-derived non-coding transcripts possess 5′-methylguanosine caps.[36] To identify the 5′ end of the intergenic transcript, we conducted RNA linker-mediated (RLM) RACE (Supplementary Figure S1).[37] We detected multiple transcription start sites (TSS) located within the 3′ untranslated region (UTR) of *C5* in Huh7 cells (Figure 2a, Supplementary Figure S1). Interestingly, transcriptional start sites in the 3′ UTR of *C5* were confirmed by Fantom5 5′cap analysis gene expression in primary hepatocytes (Supplementary Figure S2).[38] The large majority of RNAPII-derived non-coding transcripts are polyadenylated at their 3′ ends.[36]

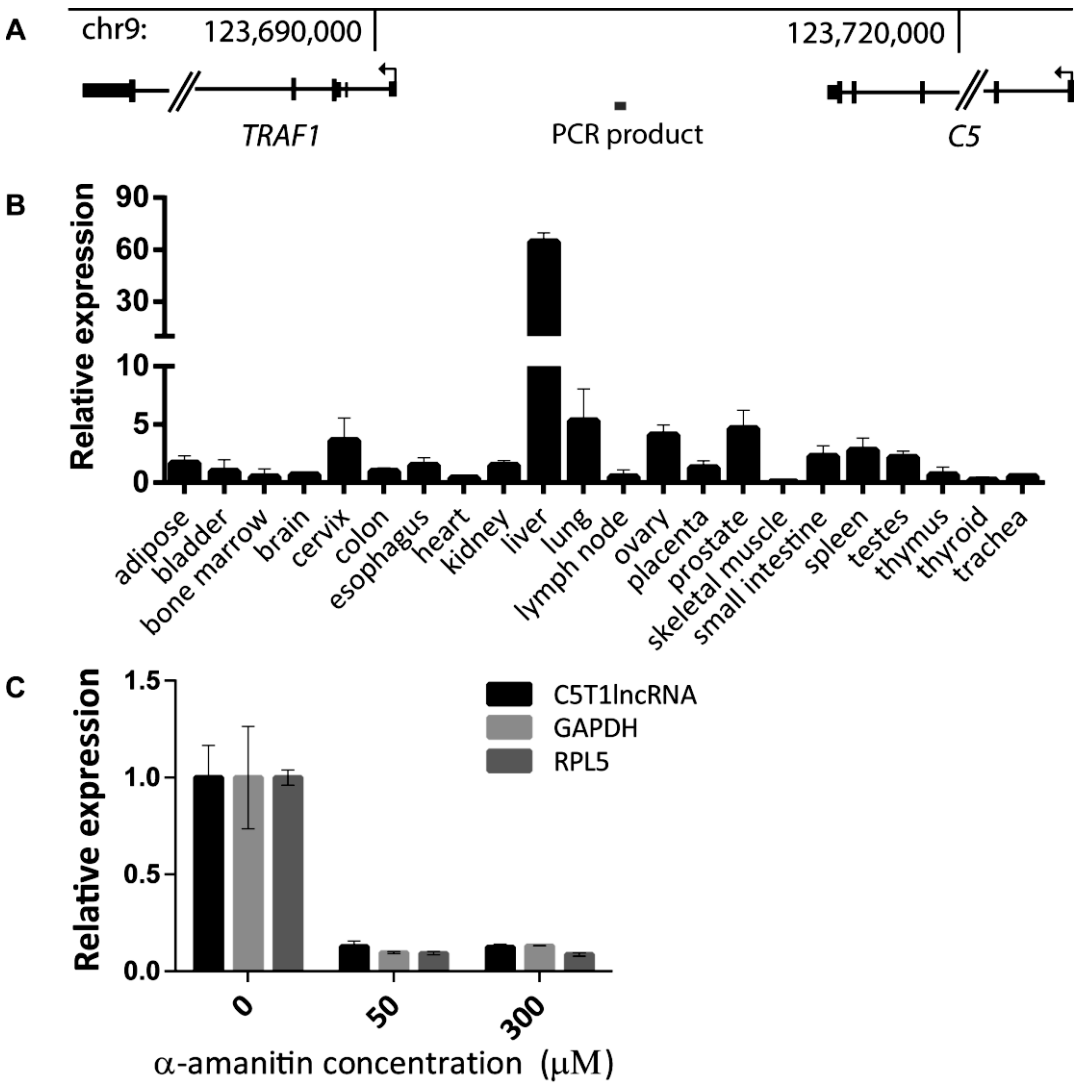

Figure 1. Identification of a novel transcript intergenic of *C5* and *TRAF1*. (a) Schematic overview of the intergenic *TRAF1-C5* region including PCR product used for RT-qPCR. Chromosome positions are according to the UCSC genome browser build GRCh37/hg19. (b) RNA levels of the intergenic amplicon (C5T1lncRNA) were measured in 22 different human tissues. Expression was normalised to GAPDH. The data are representative of two independent experiments ±s.d. (c) Inhibition of RNAPII reduced levels of the intergenic amplicon (C5T1lncRNA). A total of 50 μM α-amanitin inhibited expression of the intergenic amplicon in Huh7 cells. Expression of the intergenic transcript (C5T1lncRNA), GAPDH, RPL5 and 18S RNA was measured 36 h after addition of α-amanitin. GAPDH and RPL5 were used as positive controls, known to be transcribed by RNAPII. 18S RNA is transcribed by RNAPI and was used for normalisation. The data represents the mean±s.e.m. of two independent experiments.

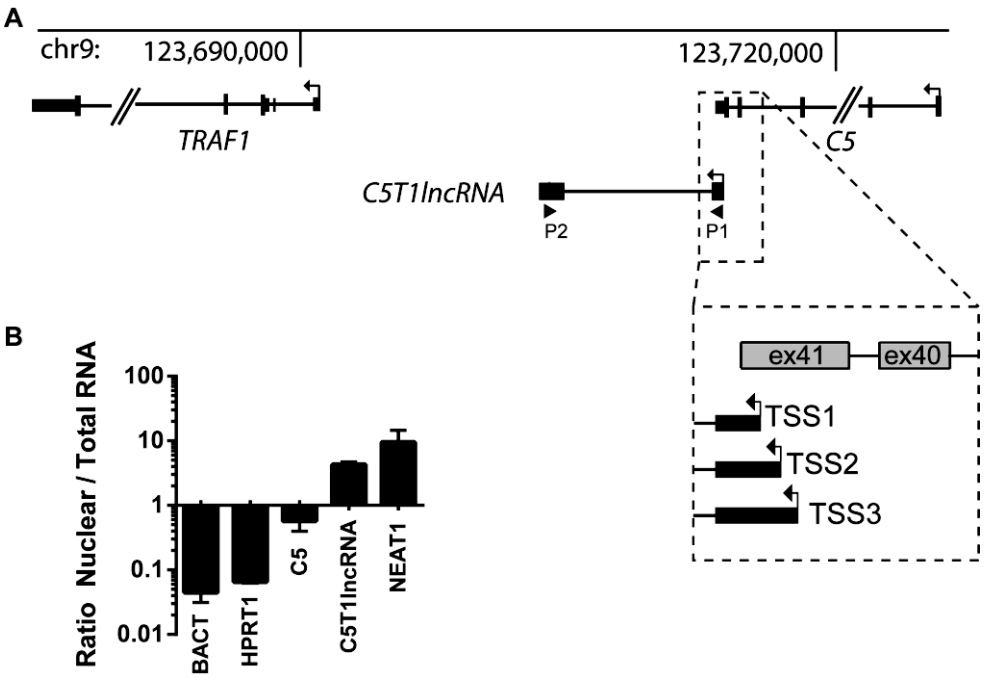

Figure 2. Characterisation of the intergenic transcript named *C5T1lncRNA*. (a) Schematic overview of an identified splice variant and the transcriptional start sites in the region harbouring the 3' UTR of *C5*. Enlargement of *C5* exon 40 and exon 41 shows the location of multiple transcriptional start sites (black blocks (TSS1, TSS2 and TSS3)). cDNA walking experiments yielded the identification of a splice variant. The spliced product was obtained by PCR on Huh7 cDNA using two primers: P1 (C5T1lncRNA5) and P2 (C5T1lncRNA2) indicated by the arrowheads. (b) Analysis of the subcellular localisation of C5T1lncRNA. Huh7 cells were separated and used for a nuclear RNA isolation and total RNA isolation, and RT-qPCR was performed to compute nuclear/total RNA ratio. qRT-PCR results are depicted as nuclear/total RNA ratio. NEAT1 is a known nuclear-localised lncRNA, whereas β-actin, HPRT1 and C5 are protein-coding mRNAs expected to localise in the cytoplasm.

3' RACE analysis in Huh7 cells revealed a 3' polyadenlation signal that does not overlap with the start of *TRAF1* (Figure 2a, Supplementary Figure S3) further confirming the presence of an independent transcript in the *TRAF1-C5* intergenic region. Further characterisation using Sanger sequencing identified two splice sites (Supplementary Figure S4).[39] The full sequence of the obtained transcript can be found in Supplementary Figure S5. Finally, we investigated whether the identified transcript is part of *C5* by a cDNA walking experiment. Primers on various exons of *C5* and the identified RNA failed to amplify a hybrid product

containing RNA from both transcripts, suggesting that the identified RNA is an independent RNA transcript (Supplementary Figure S6). We next searched for protein-coding potential of the novel transcript using NCBI open reading frame finder. All predicted open reading frames are small (<70 amino acids), do not contain known protein motifs, and more importantly have poor Kozak consensus sequences (Supplementary Figure S7a).[40,41] Moreover, the protein-coding potential score of the transcript sequence predicted by coding potential calculator (CPC) was similar to other non-coding RNAs, suggesting that it is likely of non-coding nature (Supplementary Figure S7b).[42] To find further evidence for the lncRNA nature of the identified transcript, we investigated its cellular localisation. Many lncRNAs, for example, NEAT1[43] are enriched in the nucleus.[13] We also found the identified transcripts (C5T1lncRNA) to be highly enriched in the nucleus in contrast to their neighbouring coding *C5* gene (Figure 2b). Given the presented evidence, we believe that we identified a novel lncRNA transcript in the *TRAF1-C5* region. Therefore, we refer to the transcript as *C5T1lncRNA* (*C5-TRAF1*-long non-coding RNA) from here onwards.

Correlated expression of C5T1lncRNA, TRAF1 and C5

As C5T1lncRNA is enriched in the nucleus, we hypothesised a transcriptional regulatory role for C5T1lncRNA as many regulatory lncRNAs are nuclear enriched, for example, *Pint*,[44] *Kcnq1ot1*[45] and *PANDAR*.[46] Evidence supporting a regulatory role would be either positively or negatively correlated expression of C5T1lncRNA and adjacent genes. To determine whether TRAF1 and/or C5 transcript levels correlate with C5T1lncRNA, we measured C5 and TRAF1 mRNA expression in the same tissues that were analysed for C5T1lncRNA expression (Figures 3a and b).

We only observed a strong significant positive correlation between C5 and C5T1lncRNA throughout the tissues analysed (r=0.87, P-value<0.001; Figure 3c). When separated in expression clusters, tissues belonging to cluster A predominantly coexpress *C5T1lncRNA* with *C5*. However, *C5T1lncRNA* and *TRAF1* are both expressed in spleen suggesting that perhaps *C5T1lncRNA* has a regulatory role on *TRAF1* in the cells residing there (Figure 3c, cluster D high mRNA levels of C5T1 and TRAF1). Recent data has shown that lipopolysaccharide (LPS)-induced inflammatory response in human monocytes depends on the transcription of a large number of lncRNA genes.[47]

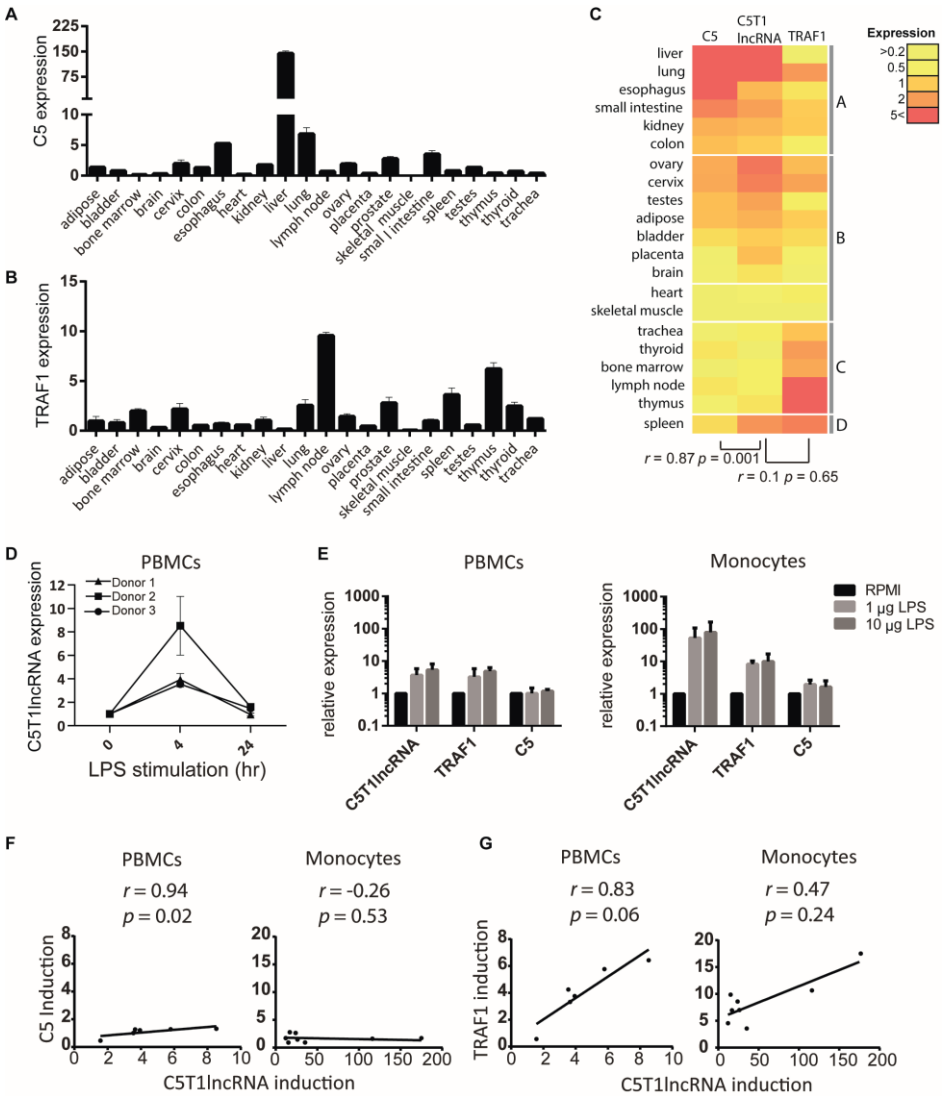

Figure 3. *C5T1lncRNA* is induced by LPS and correlates with C5 expression in various tissues and peripheral blood mononuclear cells (PBMCs). (a, b) Distribution of C5 and TRAF1 mRNA in 22 different human tissues. Expression was normalised using GAPDH. (c) Heat map of C5T1lncRNA, C5 and TRAF1 expression. Expression of each gene was compared with the mean of the 22 tissues and is indicated by colour (red=high expression/yellow=low expression). Spearman *r*-values using the expression data of all tissues are depicted underneath the heat map. Expression cluster analysis result in four clusters and are depicted as grey bars right of the heat map named A–D. (d) Time course of C5T1lncRNA expression in PBMCs after LPS stimulation (10 µg ml^{-1}). (e) Dose response curve of LPS-induced expression of C5T1lncRNA, C5 and TRAF1 in PBMCs (left) and monocytes (right). Total RNA was isolated and measured after 4 h of stimulation of either PBMCs or monocytes from three individual donors. Expression of PBMCs and monocytes were normalised for GAPDH and RPL5. C5 (f) and TRAF1 (g) RNA levels were measured within the same stimulated donors and correlated to the induction of C5T1lncRNA. Spearman *r*-values are depicted above the graphs.

As TRAF1 is abundantly expressed in monocytes[38] we analysed the expression of *C5T1lncRNA* in relation to C5 and TRAF1 transcript levels in LPS stimulated peripheral blood mononuclear cells (PBMCs) and primary monocytes. An induction of C5T1lncRNA was observed in stimulated PBMCs that was part of an early response as the highest induction was observed after 4 h (Figure 3d). A dose-dependent induction of C5T1lncRNA was observed for all tested donors and was greatest in primary monocytes (up to 200-fold for one of the donors; Figure 3e). This indicates that expression of *C5T1lncRNA* is not spurious but under specific regulation, here, being part of the innate immune response. *C5T1lncRNA* expression correlates strongly with C5 in LPS-stimulated PBMCs (r=0.94, P-value=0.02; Figures 3e and f). Interestingly, a suggestive correlation was also found between C5T1lncRNA and TRAF1 (r=0.83, P-value=0.06; Figures 3e and g). Analysis of LPS-stimulated monocytes did not show correlation between any of the genes (Figures 3f and g).

C5T1lncRNA knockdown decreases C5 levels in hepatocytes

Our data indicate co-expression of C5 with C5T1lncRNA in various tissues and PBMCs. To demonstrate that C5T1lncRNA can affect C5 mRNA levels we aimed to interfere with C5T1lncRNA levels in cells abundantly expressing the lncRNA. To this end we created C5T1lncRNA knockdowns in Huh7 cells using lentiviral vector-mediated expression of shRNAs targeting exon 2 of C5T1lncRNA. Two independent C5T1lncRNA-specific shRNAs resulted in ~60% reduction of C5T1lncRNA transcript in comparison with the shRNA control (Figures 4a and b). Knockdown of C5T1lncRNA also yielded reduced levels of C5 mRNA (Figure 4c). In contrast, C5T1lncRNA knockdown did not affect TRAF1 (Figure 4d). As the abundance of the C5T1lncRNA transcripts is an estimated 50 times lower than C5 mRNA (Supplementary Figure S8), we exclude the possibility that these results are caused by an alternative C5 transcript. Expression analysis of the generated shRNA lines demonstrated a significant correlation between C5T1lncRNA and C5 expression (r=0.85, P<0.05) (Supplementary Figure S9) corroborating our initial tissue-wide correlation data, and suggesting that C5T1lncRNA influences C5 mRNA levels.

C5T1lncRNA influences C5 RNA levels in synovial fibroblasts from RA patients

As C5T1lncRNA knockdown yielded lower C5 mRNA levels in Huh7 cells, we wished to investigate whether the observed correlated C5T1lncRNA and C5

expression can be observed in RA-relevant cell types as well. siRNA studies in primary PBMCs are not feasible, because transfection or transduction of primary blood cells is notoriously difficult. We therefore performed the knockdown experiment in synovial fibroblasts. Analysis of the mRNA in synovial fibroblasts derived from RA patients showed that *TRAF1*, *C5* and *C5T1lncRNA* are expressed (Figure 5a). The expression levels are independent of RA as synovial fibroblasts from osteoarthritis patients have similar levels of TRAF1, C5 and C5T1lncRNA (Figure 5b). As primary synovial fibroblasts grow relatively slow in culture we performed transient transfections using GapmeR sequences instead of shRNA transduction.

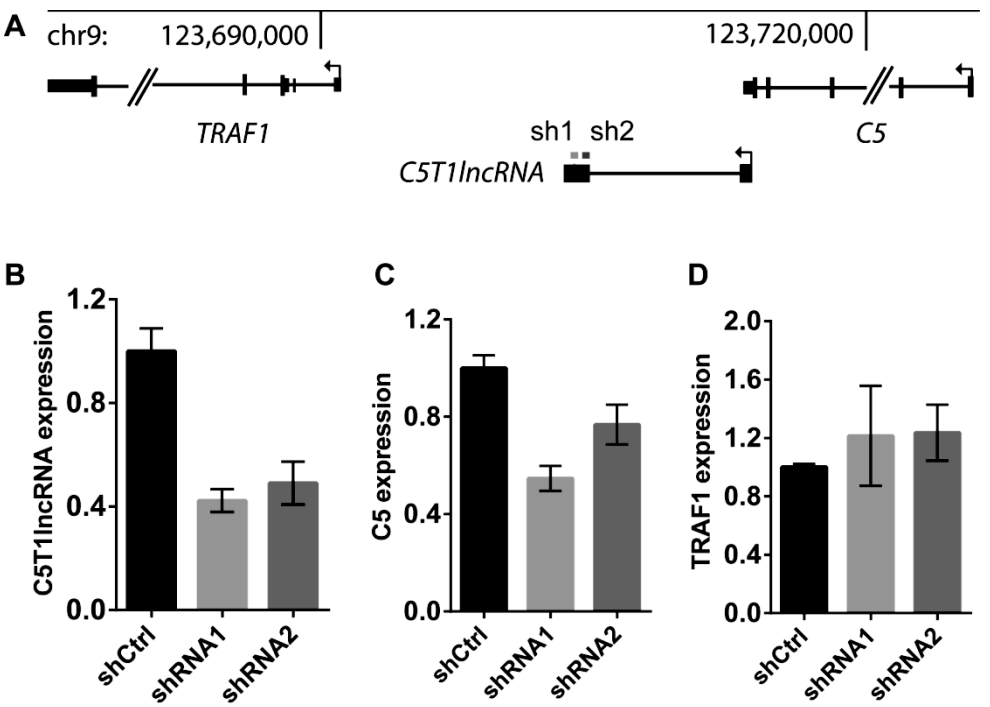

Figure 4. Knockdown of C5T1lncRNA using lentiviral shRNA. (a) Schematic representation of the location of shRNA1 (sh1) and shRNA2 (sh2) sequences that were used to knockdown C5T1lncRNA with a lentiviral vector. As a control Huh7 cells were transduced with empty viral vector and a nonspecific shRNA (shCtrl). (b-d) C5T1lncRNA (b), C5 (c) and TRAF1 (d) RNA levels were quantified using RT-qPCR. RNA levels were normalised for GAPDH. Data represent the mean±s.e.m. of two independent experiments.

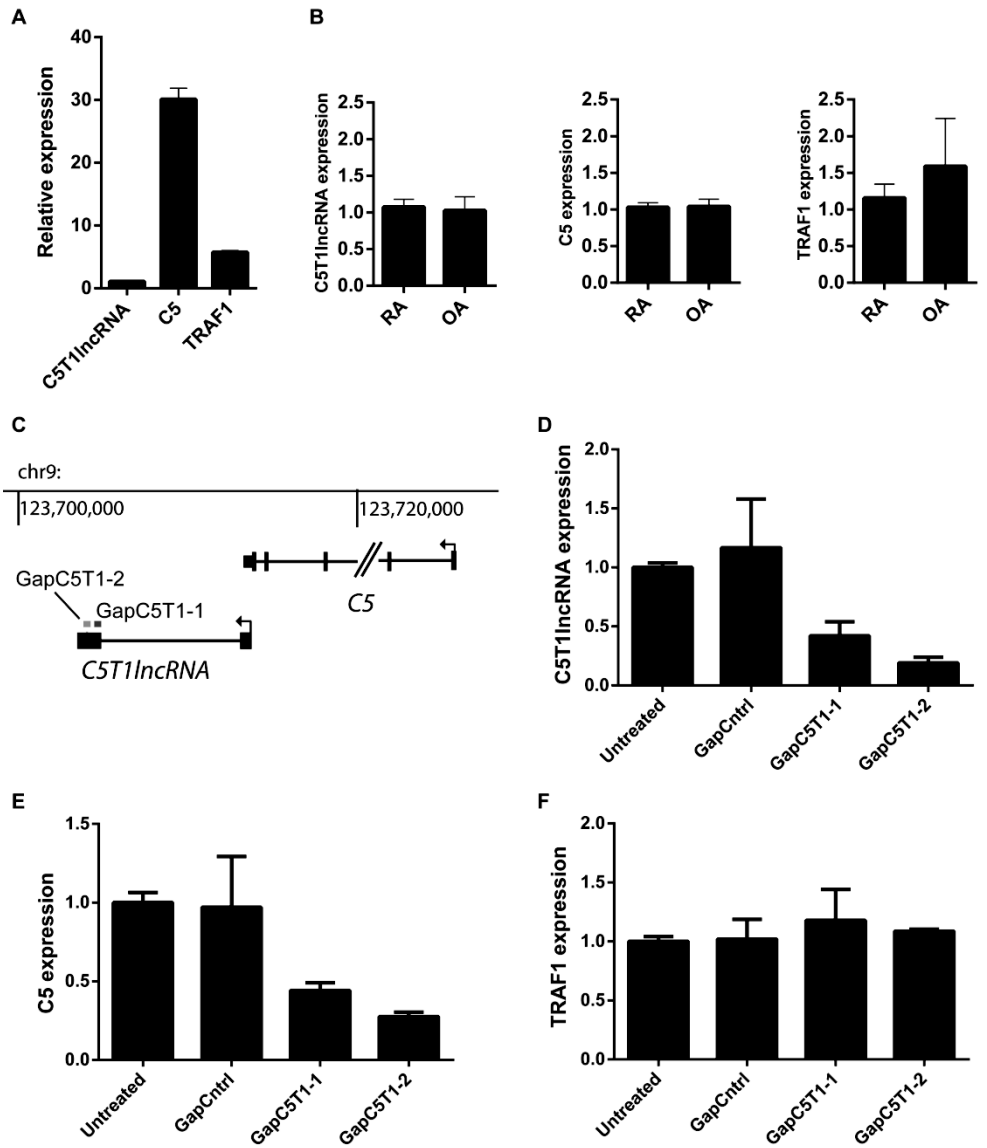

Figure 5. Expression of C5T1lncRNA, C5 and TRAF1 in synovial fibroblasts. (a) C5T1lncRNA, C5 and TRAF1 expression was measured in synovial fibroblasts obtained from two RA patients. Expression of C5T1lncRNA was set to 1. (b) C5T1lncRNA, C5 and TRAF1 RNA levels measured in synovial fibroblasts obtained from 18 RA and 10 osteoarthritis patients. Mean expression in RA patients was set to 1. RNA levels were normalised for housekeeping genes HPRT1 and B2M. (c–f) Knockdown of C5T1lncRNA using GapmeRs in synovial fibroblasts. Synovial fibroblasts were cultured untreated for 48 h or treated with a nonspecific GapmeR (GapCntrl) or two specific GapmeR sequences (GapC5T1-1 and GapC5T1-2). (c) Schematic representation of the target location of the GapmeR sequences (GapC5T1-1 and GapC5T1-2) used to knockdown C5T1lncRNA. C5T1lncRNA (d), C5 (e) and TRAF1 (f) RNA levels were quantified using RT-qPCR. RNA levels were normalised to B2M. The data represent the mean±s.e.m. of two independent donors.

GapmeR-mediated knockdown of C5T1lncRNA in synovial fibroblasts resulted in strongly reduced C5 mRNA levels reiterating the results we observed in hepatocytes (Figures 5c–e). In contrast, mRNA levels of TRAF1 were not affected by the C5T1lncRNA knockdown (Figure 5f). Taken together, we identified a novel lncRNA in a locus genetically associated to RA that influences C5 RNA levels in both liver-derived cells and synovial fibroblast.

Discussion

lncRNAs that act locally by regulating the levels of neighbouring genes are thought to be widely present in the genome. A few locally acting lncRNAs have shown to be responsible for important developmental processes and physiology, and have been linked to human disease.[48] However, locally acting lncRNAs are highly understudied as possible disease candidate genes in loci associated to complex genetic diseases. Challenges include low expression levels, higher tissue specificity compared with protein-coding genes, and transcriptional overlap (on both strands). A more prevalent role for locally acting lncRNAs in complex genetic diseases has been suggested by a study correlating lncRNA expression to disease-associated polymorphisms using PBMCs as 29 eQTLs affected expression of the lncRNAs and neighbouring protein-coding genes.[49] The outcome was however hampered by low expression levels of lncRNAs. Also more advanced RNA-sequencing methods may be hindered in the identification of lncRNA disease genes by aforementioned low lncRNA expression and complex sequence characteristics. The success of 'forward' approaches hinges on the correct choice of tissues, in which transcript levels are abundant enough to be quantitatively assessed. The aetiology of many complex genetic diseases is largely enigmatic confounding the cell type choice. We took a reverse approach based on the presence of multiple candidate genes, eQTLs, and preliminary unpublished evidence of transcription in the GWAS region. Interestingly, analysis of expression of the intergenic region between *TRAF1* and *C5* revealed the presence of a novel, spliced non-coding gene that starts in the region containing the 3' UTR of *C5*, and ends half-way between *C5* and *TRAF1*. The overlap with the region encompassing the 3' UTR of *C5*, and the low expression of C5T1lncRNA imposes that microarray or regular RNA sequencing transcriptome analyses will fail to identify this non-coding transcript. Knocking down the expression of C5T1lncRNA in cells, in which C5T1lncRNA expression correlates with C5, yielded reduced C5 mRNA levels. As

mRNA levels affected, the protein C5 is likely to be similarly decreased as well. Thus our data suggest that C5T1lncRNA influences the mRNA levels of the RA candidate gene C5, albeit we do not yet grasp the mechanism. Possible paths for the observed decreased C5 levels include miRNA sponging and transcriptional enhancement. The fact that C5T1lncRNA and C5 expression are positively correlated could point to a miRNA sponge function.[6] Sponging miRNAs is a common mechanism by which lncRNAs enhance mRNA levels of protein coding genes when the lncRNA gene overlaps with that of protein-coding genes.[6,50] To our knowledge, miRNAs targeting the 3' UTR of C5 mRNA have been predicted but not yet validated.[51] Another plausible functional mechanism would be based on the presence of a Tigger4B repeat in C5T1lncRNA. This repeat sequence is 93% identical to a stretch of 130 bp in intron 28 of *C5*. As these sequences are in the reverse orientation, the RNA could potentially hybridise to either C5 pre-mRNA or its genomic DNA. Recently, lncRNA/pre-mRNA interaction was hypothesised as a strategy used by many ncRNAs as part of their regulatory function.[52] Moreover, repeats can be found in the majority of lncRNAs and have been associated to various functional mechanisms.[53] Alu repeat sequences in lncRNAs have been linked to Staufen-mediated decay of coding transcripts,[54] Staufen-mediated mRNA stabilisation[55] and transcriptional activation.[15] Moreover, *BACE1-AS* a non-coding RNA that is transcribed from the opposite DNA strand as *BACE1* was shown to stabilise the mRNA of *BACE1* by hybridisation between both transcripts.[56] A similar regulatory mechanism might apply to *C5T1lncRNA* and *C5*, where C5 RNA is stabilised by hybridisation to C5T1lncRNA. However, the precise mechanism and its potential role in RA require further investigation. Finally, eQTL data have revealed that lower levels of C5 and TRAF1 are linked to the risk alleles. It will be of interest to see whether C5T1lncRNA expression or function is related to the disease haplotype. Remarkably, two highly RA-associated single-nucleotide polymorphisms (rs10818488 and rs35517037) are present in the exonic region of *C5T1lncRNA*, of which rs35517037 is located within the Tigger4B repeat. Such single-nucleotide polymorphisms could alter RNA function by either influencing RNA stability, RNA structure or RNA-binding ability and therefore need further investigation.[57]

Taken together, we identified a novel RA candidate gene that is of non-coding nature using an alternative approach based on the presence of protein coding candidate disease genes in genetic risk locus. This lncRNA could be considered as

a novel candidate gene as it is fully located within the associated region, is expressed in RA-relevant cells and influences the mRNA levels of at least one of the postulated candidate disease genes. The strategy employed in this study constitutes a practical alternative for the identification of novel regulatory lncRNAs as disease candidate genes of complex genetic immune-related diseases. It would be interesting to further dissect the role for this non-coding RNA in disease over the next years. It is expected that discovery of these lncRNAs will take a massive leap in near future and may represent pivotal players in the pathogenesis of complex immune diseases.

Materials and methods

Cell culture

Huh7 cells obtained from ATCC were cultured in Dulbecco's Modified Eagles Medium (Gibco, Life Technologies, Paisley, UK) supplemented with 15% fetal bovine serum (Gibco, Life Technologies), 2 mM l-glutamine (Gibco, Life Technologies), 10 U ml^{-1} penicillin–streptomycin (Gibco, Life Technologies) at 37 °C and 10% CO2. Huh7 cells were tested negative for mycoplasma multiple times during the various experiments.

Primary PBMCs isolated on a Ficoll gradient (pharmacy LUMC, Leiden, The Netherlands) were cultured in RPMI 1640 Medium (Gibco, Life Technologies) supplemented with 10% heat-inactivated fetal bovine serum, 2 mM l-glutamine and 10 U ml^{-1} penicillin–streptomycin at 37 °C and 5% CO2. Monocytes (CD14+) were isolated from PBMCs using magnetic beads (Miltenyi Biotec, Bergisch Gladbach, Germany) according to the manufacturer's instructions. The purity of isolated monocytes was determined by flow cytometry using a PerCPCy5.5 conjugated antibody against human CD14 (catalogue no. 45-0149-41: eBioscience, San Diego, CA, USA).

LPS stimulation experiments were performed with either 1 × 10^6 PBMCs or monocytes per ml with 1 or 10 µg LPS obtained from *Salmonella typhosa* (Sigma-Aldrich, Saint Louis, MO, USA). Primary cultures of synovial fibroblasts were established by collagenolytic digestion of synovial tissue specimens obtained from RA and osteoarthritis patients during joint replacement surgery. RA patients fulfilled the American College of Rheumatology 1987 criteria for RA.[58] Synovial

fibroblasts cultures were maintained in Dulbecco's Modified Eagles Medium/F-12 (Gibco, Life Technologies) supplemented with 10% fetal bovine serum, 2 mM l-glutamine, 10 mM HEPES (Gibco, Life Technologies), 50 U ml^{-1} penicillin–streptomycin, 0.5 mg ml^{-1} amphotericin B (Gibco, Life Technologies) and used for experiments between passages 5 and 8. The study was approved by the local ethic committee and all patients provided informed consent.

RNA isolation, first-strand cDNA synthesis and real-time quantitative PCR
Total RNA was extracted using nucleospin RNA purification columns (Macherey Nagel, Duren, Germany) according to the manufacturer's instructions. Purified RNA was subjected to additional DNase I (1 u μg^{-1}) (Invitrogen, Carlsbad, CA, USA) treatment. First-strand cDNA was synthesised using reverse transcriptase (Fermentas, Vilnius, Lithuania) according to the manufacturer's instructions. Real-time quantitative PCR (RT-qPCR) was conducted using sensifast no-rox SYBR green mix (Bioline, Luckenwalde, Germany) on a CFX96 or CFX384 real-time PCR system (Bio-Rad, Temse, Belgium).

The total RNA tissue panel was obtained from Ambion (RNA survey pool, Leusden, The Netherlands), whereas total RNA from human bone marrow and lymph nodes was obtained from Clontech (Saint-Germain-en-Laye, France). For these tissues 1 µg of total RNA was used for first-strand cDNA synthesis using SuperScript III reverse transcriptase (Invitrogen). For synovial fibroblast experiments, RNA was isolated using the miRNEAsy kit (Qiagen, Hilden, Germany) including on column DNAse digestion. cDNA was synthesised using MultiScribe Reverse Transcriptase (Invitrogen). FastStart Universal SYBR Green Master (Rox) (Roche, Mannheim, Germany) was used for qRT-PCR. Expression in synovial fibroblast was measured on a 7900HT Fast Real-Time PCR System (Applied Biosciences, Foster City, CA, USA). Multiple housekeeping genes were used for normalisation: glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (RNA tissue panel), GAPDH and ribosomal protein L5 (RPL5) (PBMCs and monocytes), and beta-2-microglobulin (β2M) and hypoxanthine phosphoribosyltransferase (HPRT1) (synovial fibroblasts). Relative expression was calculated using the ΔΔCT method.[59]

Expression of untreated samples was set to 1 unless otherwise stated. Sequences of the used primer pairs are listed in Supplementary Table S1. Expression of

C5T1lncRNA was measured using C5T1lncRNA primer1 and C5T1lncRNA primer2 (Supplementary Table S1). All qPCR measurements were performed in triplicate including minus RT samples.

C5T1lncRNA characterisation

5' RLM-RACE

A schematic overview of the protocol is depicted in Supplementary Figure S1. A total of 30 µg RNA from Huh7 cells was treated with DNase I (Invitrogen). A phenol/chloroform extraction was performed after each enzymatic step. RNA was treated with calf intestine alkaline phosphatase (New England Biolabs (NEB), Frankfurt, Germany). Next, RNA was treated with tobacco acid pyrophosphatase (Epicentre, Warsaw, Poland) and the RNA adapter (for sequence see Supplementary Table S1) was ligated to the liberated 5' end with RNA ligase (Thermo Scientific, Waltham, MA, USA). Strand-specific cDNA was produced using C5T1lncRNA-primer3 and amplified with an adapter-specific primer (for sequence see Supplementary Table S1) and C5T1lncRNA primer4.

PCR products were Sanger sequenced. cDNA walking experiments were performed to identify the possible splice transcripts. A splice variant was identified using primers C5T1lncRNA primer2 and C5T1lncRNA primer5. PCR products were separated on a 1% agarose gel, purified using GeneJET gel extraction kit (Thermo Scientific) and Sanger sequenced.

3' RACE

cDNA synthesis of liver RNA was performed using a 3' RACE primer (for sequence see Supplementary Table S1). C5T1lncRNA was amplified (Phusion, NEB) using the 3'RACE adapter-primer and C5T1lncRNA primer1. A second round of PCR was performed using the nested primer: C5T1lncRNA primer6. PCR products were purified from a 1% agarose gel and Sanger sequenced. Obtained sequences were aligned to the human genome (UCSC genome browser build hg19).[60]

cDNA walking experiment

A PCR was performed on cDNA from Huh7 cells using Phusion polymerase according to the manufacturer's protocol using various primers located in *C5* and *C5T1lncRNA* (C5 primer 2 to 5, C5T1lncRNA primer 2 and 5). The C5T1lncRNA

sequence is available at the DNA Data Bank of Japan (DDBJ) with accession number LC094347.

Inhibition of transcription

3×10^5 Huh7 cells were first cultured for 24 h in one well of a 24-well plate. Cells were then treated with 0, 50 and 300 µM α-amanitin (Sigma-Aldrich) for 36 h. After treatment, RNA was collected and analysed using RT-qPCR. RNA levels of GAPDH, and RPL5 (both RNAPOLII transcribed), and 18S (RNAPOLI) were analysed and used for normalisation.

Nuclear isolation, relative abundance of transcript levels and RNA localisation

Huh7 cells were isolated by incubating 5×10^6 cells in 1 ml lysis buffer I (50 mM HEPES–KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1 × cOmplete protease inhibitors (Roche) for 10 min on 4 °C roller bank. Nuclei were pelleted (1350 *g* at 4 °C for 5 min), washed carefully in lysis buffer II (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1 × cOmplete protease inhibitors (Roche) by rocking for 10 min at RT, and repelleted. Nuclei were lysed using RA1 RNA lysis buffer (Macherey-Nagel) and RNA was isolated using nucleospin columns (Macherey-Nagel). Primers representing NEAT1, a nuclear located RNA were used as positive control. To calculate the difference between relative abundance of RNA transcripts of C5 and C5T1lncRNA, cDNA from Huh7 cells was used for qPCR. To amplify C5T1lncRNA, primer4 and primer7 were used. C5 RNA was amplified using primer7 and C5-primer1 (Supplementary Table S1). Ct-values were normalised for primer efficiencies by taking along genomic DNA.

Lentiviral knockdown of C5T1lncRNA

shRNA sequences targeting C5T1lncRNA were designed, synthesised by ShineGene (Shanghai, China), and Sanger sequenced. Sequences are depicted in Supplementary Table S2. shRNA sequences were introduced into a pRRL.Super plasmid, which was generated by introducing the H1-mcs-PGKPuro part of pRetroSuper into the pRRL vector. Lentiviral vectors were generated in 293T cells using the three plasmid lentiviral production system.[61] Lentiviral titres were estimated by ELISA on basis of P24 levels (ZeptoMetrix, Buffalo, NY, USA). As controls either an empty lentiviral vector or a vector containing a shRNA sequence against mouse ATF3 were used.

Huh7 cells were transduced at a multiplicity of infection of five based on the levels of P24 (1 ng ml^{-1} of P24 equals the transduction of 2500 cells). shRNA-expressing cells were selected using 1 μg ml^{-1} Puromycin. Two weeks after puromycin selection cells were collected for RNA analysis.

GapmeR transfection

Custom designed LNA longRNA GapmeRs targeting C5T1lncRNA; 5'-GGCCTCTTCACGTAGT-3' (GapC5T1-1) and 5'-CGGGATCTGGAACATT-3' (GapC5T1-2) were purchased from Exiqon (Vedbaek, Denmark) along with the negative control LNA longRNA GapmeR - Negative Control A (catalogue no. 300610). RASF were transfected with 10 nM GapmeRs using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. After 24 h, medium was replaced and 48 h after transfection cells were lysed for RNA isolation.

Correlation and statistical analysis

Unless otherwise stated, expression experiments were performed at least three independent times or with three independent donors and is presented as mean±s.d. Statistical analysis was performed using GraphPad (GraphPad, San Diego, CA, USA). Correlation analysis was performed using spearman rank *r* correlation test. Spearman rank analysis was applied using relative expression values. Differences with *P*-values <0.05 are considered significant.

Supplementary material

Supplementary information is available online on the Genes and Immunity website (http://www.nature.com/gene): Supplementary Figure S1-S9 and Supplementary Table S1 and S2.

Acknowledgements

We wish to thank SJ Cramer and M Rabelink for the generation of shRNA-containing lentiviral particles and H van Dam for providing reagents. We also thank the volunteers for donating blood and AWG Berendsen for experimental support. Finally we thank RC Hoeben and A Zaldumbide for comments and critical reading of the manuscript. This work was supported by the Dutch Arthritis

Foundation, The Netherlands. F Kurreeman is supported via UNESCO-L'Oreal for Women in Science Fellowship, Marie Curie FP7 Outgoing Fellowship and an LUMC Research Fellowship.

References

1. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet 2011; 12: 861–874.
2. Churchill GA, Holland MJ, Draghici S, Khatri P, Eklund AC, Szallasi Z *et al*. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 2007; 316: 1484–1488.
3. Hangauer MJ, Vaughn IW, McManus MT. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. PLoS Genet 2013; 9: e1003569.
4. Van Leeuwen S, Mikkers H. Long non-coding RNAs: guardians of development. Differentiation 2010; 80: 175–183.
5. Geisler S, Coller J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat Rev Mol Cell Biol 2013; 14: 699–712.
6. Xia T, Liao Q, Jiang X, Shao Y, Xiao B, Xi Y *et al*. Long noncoding RNA associated-competing endogenous RNAs in gastric cancer. Sci Rep 2014; 4: 6088.
7. Wang X, Arai S, Song X, Reichart D, Du K, Pascual G *et al*. Induced ncRNAs allosterically modify RNA-binding proteins *in cis* to inhibit transcription. Nature 2008; 454: 126–130.
8. Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S *et al*. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol Cell 2010; 38: 662–674.
9. Terranova R, Yokobayashi S, Stadler MB, Otte AP, van Lohuizen M, Orkin SH *et al*. Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. Dev Cell 2008; 15: 668–679.
10. Tsai M, Manor O, Wan Y, Mosammaparast N, Wang JK, Shi Y *et al*. Long noncoding RNA as modular scaffold of histone modification complexes. Science 2010; 329: 689–693.
11. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J *et al*. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell 2008; 32: 232–246.
12. Nagano T, Mitchell JA, Sanz LA, Pauler FM. The air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science 2008; 322: 1717–1720.
13. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H *et al*. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 2012; 22: 1775–1789.
14. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H *et al*. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 2014; 42: D1001–D1006.
15. Holdt LM, Hoffmann S, Sass K, Langenberger D, Scholz M, Krohn K *et al*. Alu elements in ANRIL non-coding RNA at chromosome 9p21 modulate atherogenic cell functions through trans-regulation of gene networks. PLoS Genet 2013; 9: e1003588.
16. Kurreeman FaS, Padyukov L, Marques RB, Schrodi SJ, Seddighzadeh M, Stoeken-Rijsbergen G *et al*. A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. PLoS Med 2007; 4: e278.
17. Plenge RM, Seielstad M, Padyukov L, Lee T, Remmers EF, Ding B *et al*. TRAF1–C5 as a risk locus for rheumatoid arthritis — a genomewide study. NEJM 2007; 357: 1199–1209.
18. Lavorgna A, De Filippi R, Formisano S, Leonardi A. TNF receptor-associated factor 1 is a positive regulator of the NF-kappaB alternative pathway. Mol Immunol 2009; 46: 3278–3282.
19. Ma X, Xu S. TNF inhibitor therapy for rheumatoid arthritis. Biomed Rep 2013; 1: 177–184.

20. Cooke TD, Hurd ER, Jasin HE, Bienenstock J, Ziff M. Identification of immunoglobulins and complement in rheumatoid articular collagenous tissues. Arthritis Rheum 1975; 18: 541–551.
21. Wang Y, Kristan J, Hao L, Lenkoski CS, Shen Y, Matis LA. A role for complement in antibody-mediated inflammation: C5-deficient DBA/1 mice are resistant to collagen-induced arthritis. J Immunol 2000; 164: 4340–4347.
22. Dunkelberger JR, Song W-C. Complement and its role in innate and adaptive immune responses. Cell Res 2010; 20: 34–50.
23. Su X, Li S, Meng M, Qian W, Xie W, Chen D *et al*. TNF receptor-associated factor-1 (TRAF1) negatively regulates Toll/IL-1 receptor domain-containing adaptor inducing IFN-beta (TRIF)-mediated signaling. Eur J Immunol 2006; 36: 199–206.
24. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS *et al*. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci USA 2009; 106: 9362–9367.
25. Messemaker T, Toes REM, Kurreeman F, Mikkers HMM. Comment on "functional analysis of a complement polymorphism (rs17611) associated with rheumatoid arthritis. J Immunol 2015; 195: 3–4.
26. Ding J, Eyre S, Worthington J. Genetics of RA susceptibility, what comes next? RMD Open 2015; 1: e000028–e000028.
27. Messemaker TC, Huizinga TW, Kurreeman F. Immunogenetics of rheumatoid arthritis: understanding functional implications. J Autoimmun 2015; 64: 74–81.
28. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K *et al*. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2013; 506: 376–381.
29. Volders P-J, Helsens K, Wang X, Menten B, Martens L, Gevaert K *et al*. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res 2013; 41: D246–D251.
30. Chakraborty S, Deb A, Maji RK, Saha S, Ghosh Z. LncRBase: an enriched resource for lncRNA information. PLoS One 2014; 9: e108010.
31. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-vega B, Regev A *et al*. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 2011; 25: 1915–1927.
32. Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddeloh JA *et al*. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nat Biotechnol 2012; 30: 99–104.
33. Warfel AH. Specific Inhibition of nuclear RNA polymerase II by a-Amanitin. Science 1970; 170: 447–449.
34. Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J *et al*. Widespread transcription at neuronal activity-regulated enhancers. Nature 2010; 465: 182–187.
35. Ren B. Transcription: enhancers make non-coding RNA. Nature 2010; 465: 173–174.
36. Tuck AC, Tollervey D. A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. Cell 2013; 154: 996–1009.
37. Liu X, Gorovsky MA. Mapping the 5' and 3' ends of Tetrahymena therrmophila mRNAs using RNA ligase mediated amplification of cDNA ends (RLM-RACE). Nucleic Acids Res 1993; 21: 4954–4960.
38. Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S *et al*. Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol 2015; 16: 22.
39. Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human splicing finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 2009; 37: e67.
40. Kozak M, Limited IRLP. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 1987; 15: 8125–8148.

41. Nishikawa T, Ota T, Isogai T. Prediction whether a human cDNA sequence contains initiation codon by combining statistical information and similarity with protein sequences. Bioinformatics 2000; 16: 960–967.
42. Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L *et al*. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res 2007; 35: W345–W349.
43. Imamura K, Imamachi N, Akizuki G, Kumakura M, Kawaguchi A, Nagata K *et al*. Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. Mol Cell 2014; 53: 393–406.
44. Marín-Béjar O, Marchese FP, Athie A, Sánchez Y, González J, Segura V *et al*. Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2. Genome Biol 2013; 14: R104.
45. Lee JT, Bartolomei MS. X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell 2013; 152: 1308–1323.
46. Han L, Zhang E, Yin D, Kong R, Xu T, Chen W *et al*. Low expression of long noncoding RNA PANDAR predicts a poor prognosis of non-small cell lung cancer and affects cell apoptosis by regulating Bcl-2. Cell Death Dis 2015; 6: e1665.
47. IIott NE, Heward JA, Roux B, Tsitsiou E, Fenwick PS, Lenzi L *et al*. Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes. Nat Commun 2014; 5: 3979.
48. Chen G, Wang Z, Wang D, Qiu C, Liu M, Chen X *et al*. LncRNADisease: a database for long-non-coding RNA-associated diseases. Nucleic Acids Res 2013; 41: D983–D986.
49. Kumar V, Westra H-J, Karjalainen J, Zhernakova DV, Esko T, Hrdlickova B *et al*. Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. PLoS Genet 2013; 9: e1003201.
50. Kallen AN, Zhou X-B, Xu J, Qiao C, Ma J, Yan L *et al*. The imprinted H19 LncRNA antagonizes Let-7 microRNAs. Mol Cell 2013; 52: 101–112.
51. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005; 120: 15–20.
52. Engreitz JM, Sirokman K, McDonel P, Shishkin AA, Surka C, Russell P *et al*. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. Cell 2014; 159: 188–199.
53. Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G *et al*. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet 2013; 9: e1003470.
54. Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature 2011; 470: 284–288.
55. Kretz M, Siprashvili Z, Chu C, Webster DE, Zehnder A, Qu K *et al*. Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature 2013; 493: 231–235.
56. Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE *et al*. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of B-secretase expression. Nat Med 2008; 14: 723–730.
57. Naslavsky MS, Crovella S, Lima Filho JL, Rocha CRC. The sound of silence: human beta-defensin-1 gene untranslated SNPs change the predicted mRNA secondary structure in a length-dependent manner. Immunol Lett 2010; 129: 53–55.
58. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988; 31: 315–324.
59. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001; 25: 402–408.

60. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M *et al*. The UCSC Genome Browser database: 2015 update. Nucleic Acids Res 2015; 43 (Database issue): D670–D681.

61. Bai Y, Soda Y, Izawa K, Tanabe T, Kang X, Tojo A *et al*. Effective transduction and stable transgene expression in human blood cells by a third-generation lentiviral vector. Gene Ther 2003; 10: 1446–1457.

Chapter 6

Antisense long non-coding RNAs are deregulated in skin tissue of patients with systemic sclerosis

J Invest Dermatol. 2017 Nov 24. pii: S0022-202X(17)33169-X. doi: 10.1016/j.jid.2017.09.053.

Tobias C. Messemaker, Loubna Chadli, Guoshuai Cai , Varshna S. Goelela, Maaike Boonstra, Annemarie L. Dorjée, Stefan N. Andersen, Harald M. M. Mikkers, Peter van ‘t Hof, Hailiang Mei, Oliver Distler, Harmen H. M. Draisma, Michael Johnson, Nicole Orzechowski, Robert W Simms, Rene E. M. Toes, Jamil Aarbiou, Tom W. Huizinga, Michael L. Whitfield, Jeroen DeGroot, Jeska de Vries-Bouwstra, Fina Kurreeman

ABSTRACT

Systemic sclerosis (SSc) is an autoimmune disease characterized by fibrosis of skin and multiple organs of which the pathogenesis is poorly understood. Here we studied differentially expressed coding and non-coding genes in relation to SSc pathogenesis with a specific focus on antisense non-coding RNAs. Skin biopsy-derived RNAs from fourteen early SSc patients and six healthy individuals were sequenced with ion-torrent and analysed using DEseq2.

Overall, 4901 genes with a fold change >1.5 and a false discovery rate < 5% were detected in patients versus controls. Upregulated genes clustered in immunological, cell adhesion and keratin-related processes. Interestingly, 676 deregulated non-coding genes were detected, 257 of which were classified as antisense genes. Sense genes expressed opposite of these antisense genes were also deregulated in 42% of the observed sense-antisense gene pairs. The majority of the antisense genes had a similar effect sizes in an independent North American dataset with three genes (CTBP1-AS2, OTUD6B-AS1 and AGAP2-AS1) exceeding the study-wide Bonferroni-corrected ρ-value (P_{Bonf}<0.0023, $P_{combined}$ = $1.1x10^{-9}$, $1.4x10^{-8}$, $1.7x10^{-6}$, respectively). In this study, we highlight that together with coding genes, (antisense) long non-coding RNAs are deregulated in skin tissue of SSc patients suggesting a novel class of genes involved in pathogenesis of SSc.

INTRODUCTION

Systemic sclerosis (SSc) is a heterogeneous complex autoimmune disease affecting connective tissues. Its pathogenesis remains elusive, but patients harbour vascular changes like Raynaud's phenomenon, autoimmunity with the presence of distinct autoantibodies, activation of both innate and adaptive immunity and active deposition of extracellular matrix leading to fibrosis. Progression of vascular and fibrotic organ damage accounts for a large proportion of the chronic morbidity and mortality up to 25% in the first five years after diagnosis in SSc (Rubio-Rivas *et al*. 2014).

In order to further understand the processes involved in SSc pathophysiology, several groups have performed gene expression studies in peripheral blood and skin of SSc patients (Gardner *et al*. 2006; Pendergrass *et al*. 2012; Milano *et al*. 2008; Whitfield *et al*. 2003). These studies have revealed that

gene expression profiles in skin from SSc patients not only differ from healthy skin but are associated with skin disease severity (Milano *et al*. 2008). Interestingly, several SSc-specific gene sets have been identified which include fibrosis related pathways involved in skin thickening (TGF-β related genes, collagen genes) as well as immunological and keratin-related pathways (interferon genes, activated macrophage genes, chemokine-related genes and keratin genes) (Mahoney *et al*. 2015; Assassi *et al*. 2015; Gardner *et al*. 2006; Mathes *et al*. 2014). These studies were all performed using microarrays, and focussed on the identification of protein coding genes and pathways that are differently regulated in SSc, and as a consequence missing an important component of non-coding genes involved in disease pathogenesis. With the use of next generation sequencing, transcriptomics studies can now shed light on the non-coding genome and the role of long non-coding RNAs (lncRNAs) in disease mechanisms.

lncRNAs represent an important layer of genome regulation and their role in the context of SSc is currently unknown. lncRNAs are transcripts over 200 nucleotides in length and come in diverse flavours including: antisense RNAs, long intergenic non-coding RNAs (lincRNAs) and pseudogenes (Derrien *et al*. 2012). Although the function of the majority of lncRNAs remains unknown, a role in regulating and shaping the genome has been proposed (Rinn JL 2013; Melé and Rinn 2016). Specifically, antisense RNAs can influence RNA levels of their sense counterpart (Faghihi and Wahlestedt 2010; Derrien *et al*. 2012; Chan *et al*. 2015; Peng *et al*. 2015; Kimura *et al*. 2013). In diseases like SSc, where deregulated gene expression signatures are present, identification of such regulatory genes may represent interesting candidates as biomarkers or unlock novel treatment avenues. In addition, compared to coding genes, lncRNAs display higher tissue specificity in their expression patterns (Derrien *et al*. 2012). Recently, deregulated lncRNA expression has been described in the skin of patients with psoriasis (Gupta *et al*. 2016) and in the regulation of TGF-β mediated processes (Richards *et al*. 2015) suggesting that lncRNAs may also be deregulated in skin of SSc patients.

In order to extend the current knowledge of the gene expression signature in SSc, we have performed RNA sequencing on skin biopsies of SSc patients and healthy controls and investigated deregulated expression of both coding and non-coding genes. Moreover, main findings on non-coding genes were replicated in an independent dataset.

RESULTS

DE genes in SSc patients are enriched in immunological, cell activation and keratinization pathways and overlap with previous studies.

In order to identify genes and pathways involved in SSc pathophysiology, we evaluated RNA expression levels in patients and controls. 4901 genes were DE with a minimum fold change of 1.5 and FDR *p*-value below 0.05 (Supplementary File 2). Hierarchical clustering on basis of these DE genes separates patients from healthy controls with the exception of 1 patient which displays a normal-like expression pattern (Supplementary Figure 1). Pathway analysis of overexpressed genes shows an enrichment in the immune response, cell activation and keratinization pathways (Supplementary File 3). Cross comparison with DE genes from a recent publication by Assassi *et al* indicates a small highly consistent (>96%) overlap with the most prominent common pathways belonging to the immunological and cell adhesion related processes (Figure 1a-c, Supplementary File 4).

In-depth analysis of specific SSc-related gene sets highlights additional candidate genes implicated in SSc and an inflammatory gene signature.

As an initial approach, we performed an in-depth analysis of several SSc gene sets which previously came forward from microarray studies including TGFβ signalling, collagen, keratin, interferon, alternative macrophage activation genes and chemokines (Figure 2, Supplementary File 1 and Supplementary File 5).

Similar to our GO-term enrichment analysis, a clear increased TGFβ expression profile that is involved in many fibrotic processes was not observed in our patient population as only 5 out of 86 TGFβ signalling genes were significantly increased (Figure 2a). On the other hand, TGFβ-gene COMP was found increased in patients as similar to previous reports (Farina *et al*. 2009; Assassi *et al*. 2015; Gardner *et al*. 2006). Moreover, many collagen and keratin associated genes are significantly increased in patients (Figure 2b and c). Also, 33 out of 97 genes from the interferon and macrophage gene sets were significantly increased in SSc patients (Figure 2d and e) indicating an increased inflammatory gene signature being present in early SSc patients (Assassi *et al*. 2015; Greenblatt *et al*. 2012; Mahoney *et al*. 2015). This observation is in line with previous studies showing that in early SSc (as is our population) the inflammatory signature is more prevalent (Assassi *et al*. 2015).

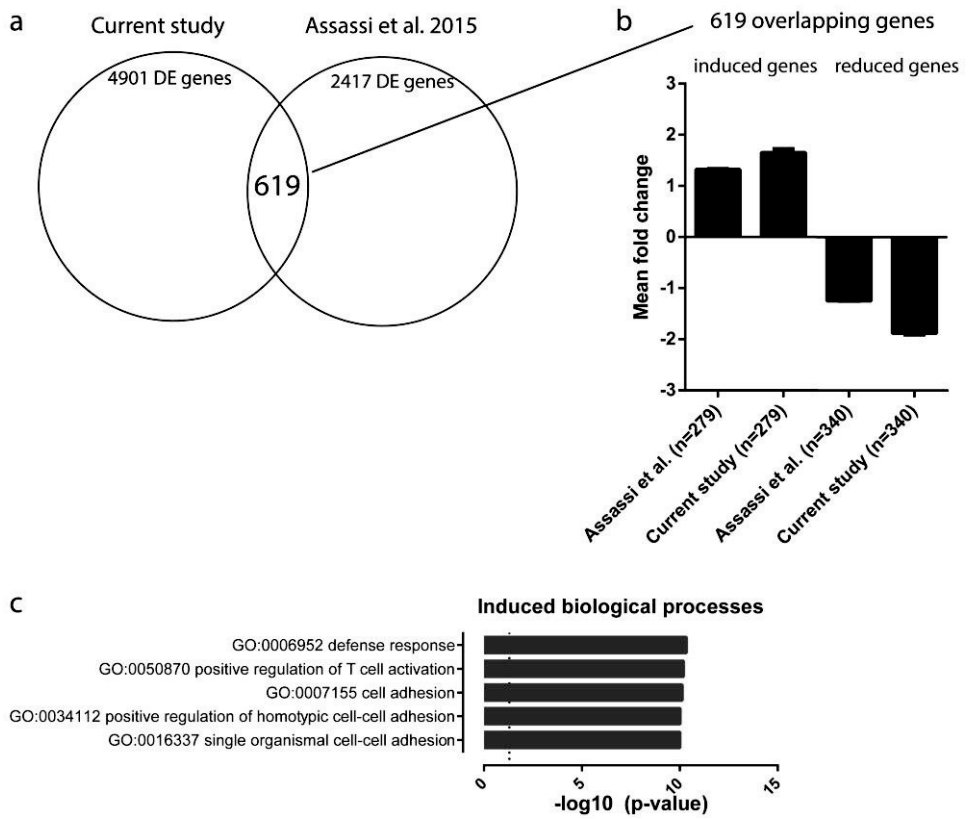

Figure 1. DE genes overlap with a previous microarray study and reveals consistent deregulated pathways. (a) Venn diagram comparing DE genes in SSc patients versus controls from the current study (n = 4901 DE genes) with a microarray study from Assassi *et al*. (n = 2417 DE genes). (b) Directionality of 619 consistently deregulated genes from the two studies displayed as mean fold change (mean ± SE). Genes up or down regulated from Assassi *et al* were selected and plotted. The concomitant fold changes of these genes from our study were also plotted indicating similar directionality in both studies. (c) Top 5 Biological processes GO-terms enriched using genes that are upregulated in SSc patients from the two studies.

Since skin paraffin sections were available for the patients under study, we stained skin sections for CD68, a marker for macrophages. In line with the observed inflammatory gene signature, clusters of macrophages were detected in the skin of SSc patients (Supplementary Figure 2). Besides these observations, several (to our knowledge previously unreported) genes including COL4A4, Keratin 4 and 9, TNFAIP3, CX3CR1, CXCL2 and PF4 were strongly deregulated in SSc patients (Figure 2b-k, Supplementary Table 1).

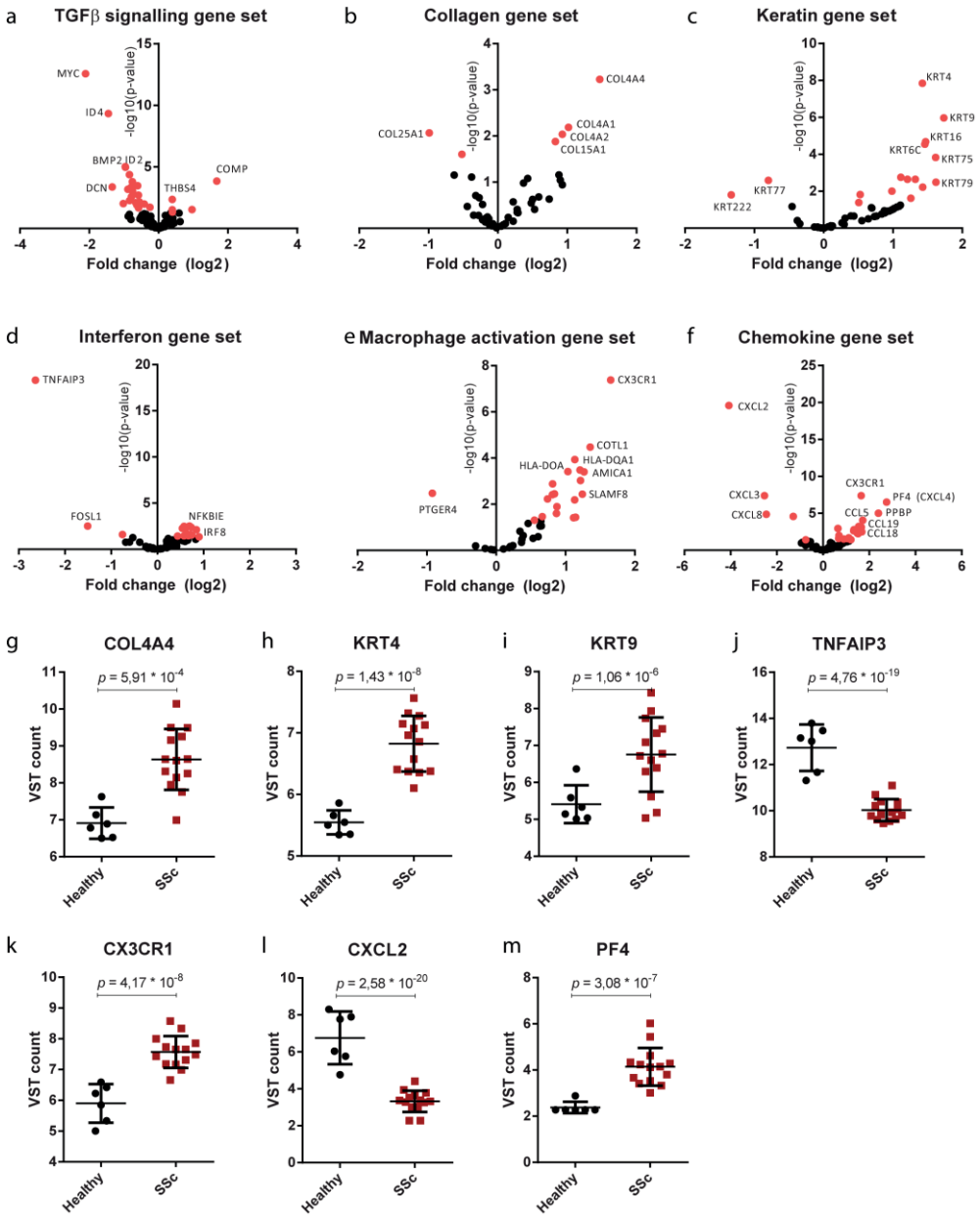

Figure 2. Analysis of DE genes of specific SSc gene sets. Volcano plots showing differential expression within the 6 genesets: TGFβ signalling (a, n=86 genes), collagen (b, n=46 genes), keratin (c, n=76 genes), interferon (d, n=50 genes), alternative macrophage activation (e, n=60 genes) and chemokine (f, n=84 genes). Genes depicted in red were significantly deregulated (Benjamini Hochberg-corrected *p*-value < 0.05). RNA levels (VST count) of individuals genes in healthy controls and SSc patients for COL4A4 (g), KRT4 (h), KRT9 (i), TNFAIP3 (j), CX3CR1 (k), CXCL2 (l), and PF4 (m). *p*-values represent Benjamini-Hochberg-corrected *p*-values. The mean ±SD of each group is depicted in the graphs.

Identification of DE lncRNAs in SSc skin biopsies in comparison to healthy controls.
In addition to coding genes, RNA sequencing allows the query of non-coding genes. Among 15941 annotated lncRNAs, 4171 were expressed in our skin biopsies. 676 lncRNAs were DE (FDR < 0.05) between SSc patients and healthy controls and show a clear differential expression signature (Figure 3a).

All 676 DE lncRNAs are listed in Supplementary File 6. Out of 676 lncRNA genes, 122 genes were decreased, while the expression of 554 genes was increased in SSc patients as compared to healthy controls.Interestingly, clustering analysis using different selection criteria of lncRNAs all displayed a pattern where non-clinically active patients clustered within the patient population and separate from controls (Supplementary Figure 3). In total, 348 lncRNAs displayed over 2 fold differential expression and the top upregulated gene is CAPN10-AS1, an antisense lncRNA (Figure 3b). Interestingly, among the 676 deregulated lncRNAs, the largest proportion (38%) belongs to the antisense gene category (Figure 3c). nAntisense lncRNAs have recently been described to have important regulatory roles on their coding gene counterparts expressed in the sense direction (Pelechano and Steinmetz 2013; Werner 2013; Katayama *et al*. 2005; Villegas and Zaphiropoulos 2015). The relevance of the antisense genes in our data set was therefore investigated.

Identification of DE antisense genes in SSc patients and their link to sense coding genes.
In order to gain further insight into the possible role of antisense RNAs in SSc, we focused our analysis on antisense genes of which a sense gene was annotated (also known as sense-antisense (SAS) gene pairs). Close proximity of antisense genes with sense genes have been linked to co-expression and co-regulation within such a SAS gene pair (Villegas and Zaphiropoulos 2015; Katayama *et al*. 2005). Out of 257 DE antisense genes, 62 have an annotated sense gene. Interestingly, an important proportion (26 out of 62) of these SAS gene pairs includes both a significant DE antisense gene and a significant DE sense gene (FDR < 0.05) (Figure 3d). We further explored the relation between sense and antisense genes using correlation analysis by comparing the correlation of gene pairs where both genes are deregulated compared to gene pairs which were not deregulated in patients (consisting of gene pairs of which only one of the two genes was deregulated and of gene pairs of which neither the sense gene nor the

antisense gene was deregulated in patients). Here high correlations (median r > 0.7) were observed for gene pairs significantly deregulated in SSc (SSc gene pairs) and were significantly higher in comparison with non SSc-deregulated gene pairs ($P < 0.001$) (Figure 3e).

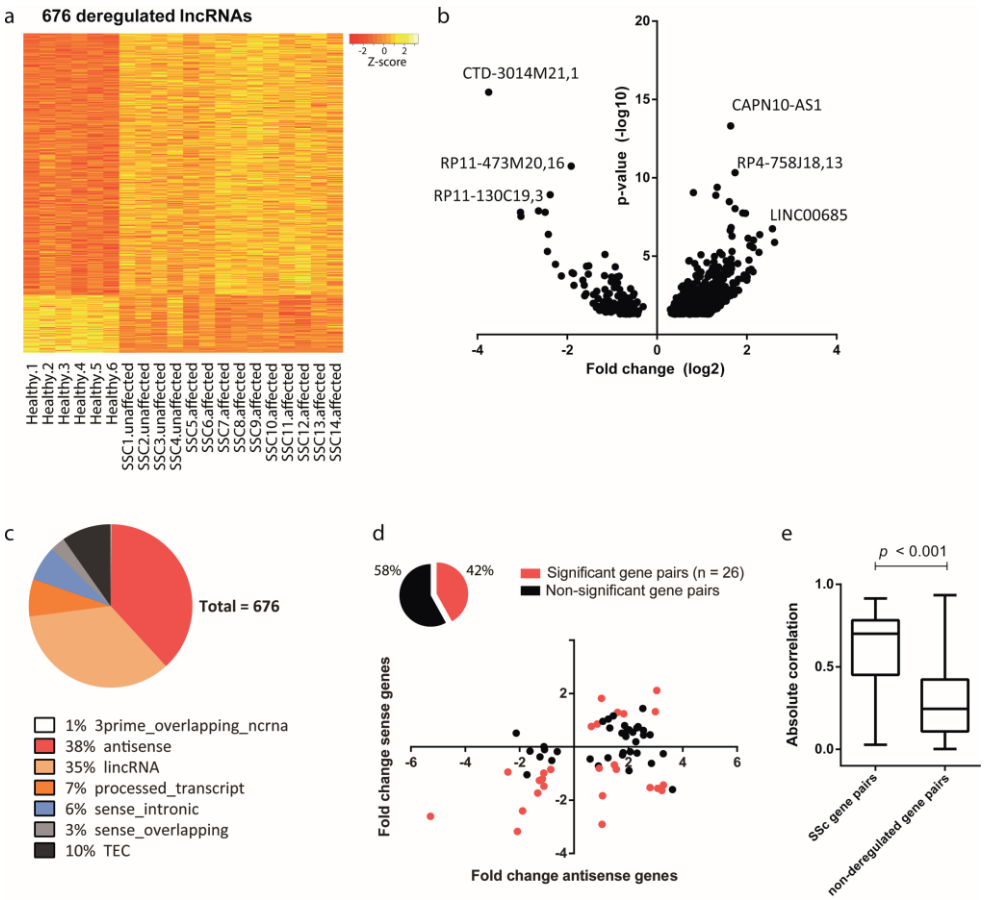

Figure 3. DE lncRNAs in SSc patients in comparison with healthy controls. (a) Heatmap depicting the Z-scores of 676 deregulated lncRNAs. Red colour indicates low expression and the yellow colour indicates high expression. (b) Volcano plot showing top deregulated lncRNAs by fold change (log2) on the x-axis and the *p*-value (-log10) on the y-axis. (c) Deregulated lncRNAs (n = 676) divided by subclasses. (d) Venn diagram and scatter plot showing the proportion of significant gene pairs (Benjamini Hochberg-corrected *p*-value < 0.05). Significant DE gene pairs are depicted in red and depicting the fold change (log2) of both the sense and antisense genes. e, Absolute spearman rank correlation between sense and antisense genes within SSc gene pairs and gene pairs not deregulated in SSc.

These data indicate that the identified antisense genes are either coexpressed with coding genes or involved in the regulation of their levels, illustrating a mechanism by which long non-coding (antisense) RNAs may play a role in SSc.

In order to obtain further evidence for the involvement in SSc of the selected 26 antisense genes, we acquired gene expression values from an independent dataset where RNA sequencing had been performed (14 SSc patients, 6 controls, Whitfield *et al*, unpublished data). 4 of the 26 genes were not present due to low expression in the independent dataset and were further excluded from the analysis. 12 out of 22 genes follow the same direction of association in both datasets (Table 1).

Table 1. Replication of 22 antisense genes in an independent RNA-seq dataset. The table includes, Fold changes (Log2FC) and *p*-values (P) from both studies and a combined *p*-value. Combined *p*-values were not calculated for the genes with opposite direction of association according to Rau *et al*. 2014 (Rau, Marot, and Jaffrézic 2014).

	Dataset 1			Dataset 2		
Gene	Log2FC	P	FDR	log2FC	P	Combined
CTBP1-AS2	0,32	0,012	0,044	0,40	7,5E-07	1,1E-09
OTUD6B-AS1	-0,95	7,0E-05	0,001	-0,63	0,001	1,4E-08
AGAP2-AS1	0,50	0,006	0,027	0,34	0,002	1,7E-06
HAND2-AS1	-1,04	0,002	0,011	-0,63	0,007	2,1E-06
HMGN3-AS1	-0,64	0,009	0,034	-0,33	0,017	2,6E-05
ZBTB11-AS1	0,53	0,002	0,010	0,17	0,143	4,5E-05
NIFK-AS1	-0,56	0,006	0,027	-0,36	0,178	2,3E-04
WAC-AS1	-0,58	0,001	0,009	-0,17	0,217	5,9E-05
PIK3CD-AS2	1,50	5,1E-06	1,5E-04	0,18	0,407	3,1E-07
ARRDC1-AS1	0,43	0,012	0,045	0,13	0,411	0,001
ZNF252P-AS1	1,64	1,1E-04	0,001	0,19	0,422	7,8E-06
SBF2-AS1	-0,43	0,014	0,049	-0,06	0,715	0,002
UNC5B-AS1	1,52	5,6E-05	0,001	-0,76	0,007	NA
HOXA10-AS	-2,64	4,4E-11	1,3E-08	0,53	0,056	NA
SLC25A25-AS1	0,52	4,3E-04	0,004	-0,30	0,163	NA
RUNDC3A-AS1	0,92	0,001	0,005	-0,30	0,225	NA
ZBED5-AS1	0,45	0,012	0,044	-0,16	0,275	NA
LOXL1-AS1	0,80	5,6E-05	0,001	-0,22	0,408	NA
BRWD1-AS2	1,54	3,4E-07	2,0E-05	-0,14	0,514	NA
ZEB1-AS1	-0,67	0,003	0,015	0,06	0,738	NA
RGMB-AS1	0,78	0,005	0,023	-0,05	0,815	NA
TMPO-AS1	1,40	7,4E-08	5,8E-06	-0,02	0,923	NA

Three antisense genes CTBP1-AS2, OTUD6B-AS1 and AGAP2-AS1 reached beyond the study-wide replication *p*-value threshold (P < 0.0023) (Table 1 and Figure 4a-c). Verification using an second experimental approach confirmed that these three genes are significantly deregulated ((P < 0.01), Supplementary Figure 4).

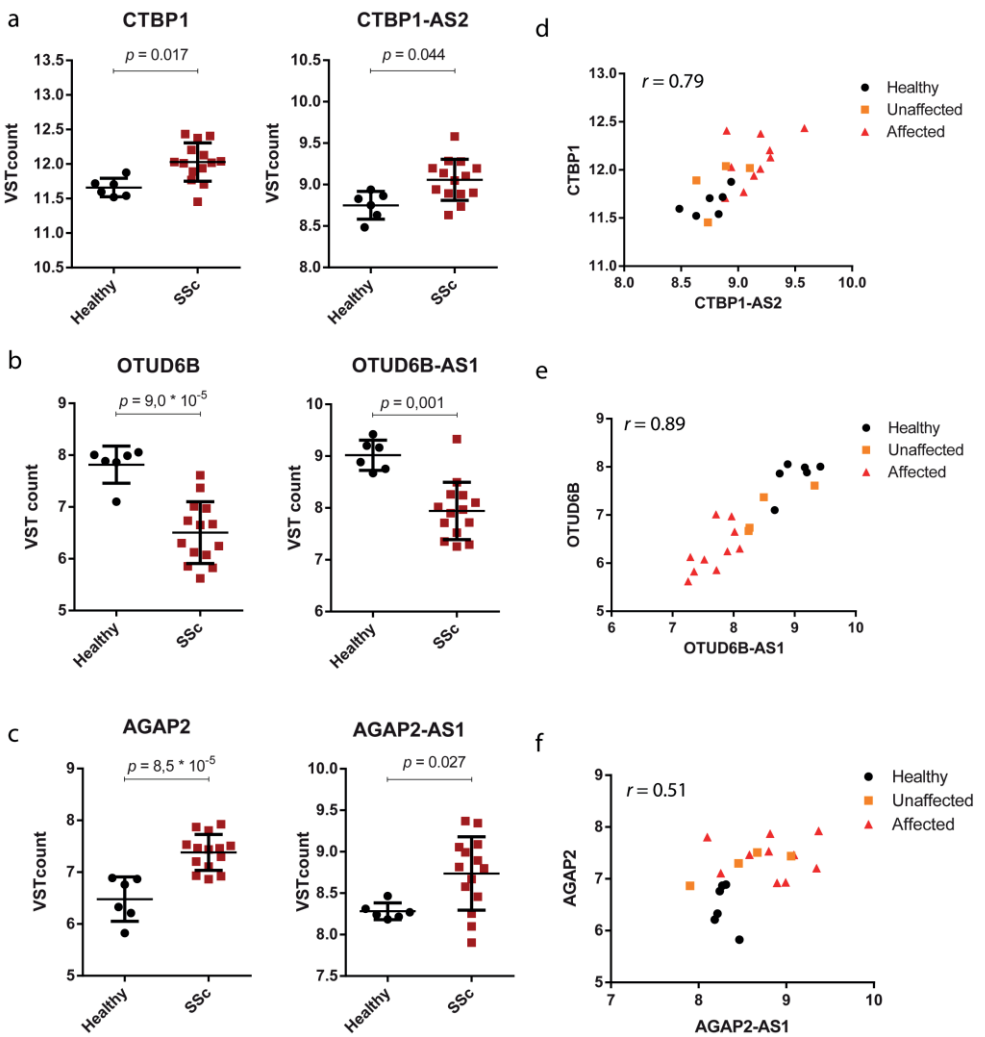

Figure 4. Top 3 replicated antisense genes show strong correlation with their sense coding gene. (a-c) VST count values of top 3 replicated SAS gene pairs: CTBP1 (a), OTUD6B (b) and AGAP2 (c) in SSc patients (n = 14) and controls (n = 6) *p*-values are Benjamini-Hochberg corrected and were generated via DEseq2. (d-f) Correlation between sense and antisense genes within a gene pair for CTBP1 (d), OTUD6B (e) and AGAP2 (f). Count values are divided into healthy, unaffected or affected skin tissue. Spearman rank test was used to calculate correlations between the sense and antisense gene.

We confirmed the non-coding nature of these antisense genes using a coding potential calculator which showed an overall low coding potential for CTBP1-AS2, OTUD6B-AS1 and AGAP2-AS1 (Supplementary Figure 5). We next evaluated the relationship of these non-coding antisense genes with their paired sense gene across our patients and controls. Interestingly, the identified antisense genes show a strong correlation with their paired sense gene across the 20 individuals, in particular for OTUD6B-AS1 and CTBP1-AS2 (r = 0.89, P < 0.001 and r = 0.79, P < 0.001, respectively, Figure 4D-F). As skin is composed of many cell types we took advantage of available cell type specific expression datasets to gain further insight into which cell types may be relevant for these candidates. CTBP1 and CTBP1-AS2 levels also positively correlate across specific cell types and this correlation is highest in immune cells (r = 0.7, P <0.001) (Figure 5a). The OTUD6B gene pair is expressed in dermal and immune cells, and shows a correlation that was similar as observed across patients (r = 0.6-0.8, P<0.01) (Figure 5b).

Interestingly, AGAP2 is only expressed in immune cells while AGAP2-AS1 is only expressed in dermal cell types (Figure 5c). Finally, we further investigated the correlation of these gene pairs in the replication dataset. These data show that the CTBP1 and OTUD6B gene pairs also display a significant correlation (r > 0.8, P <0.001 for both gene pairs) in the replication dataset (Supplementary Figure 6) while the correlation for AGAP2 is absent in the replication dataset (r = 0.21). These results seem to coincide with the tissue-specific expression data obtained from FANTOM5 were a positive correlation between AGAP2-AS1 and AGAP2 is also absent. Altogether, we identified non-coding genes that are expressed in cell-types relevant for SSc and of which the levels are altered in a disease specific manner in the skin of SSc patients.

DISCUSSION

Our results using next generation sequencing firstly confirmed previous studies using microarrays and confirmed an inflammatory signature in the skin of early SSC patients. In addition to the analyses on coding genes, we report an in-depth analysis of deregulated lncRNAs in skin tissue from SSc patients. The top-3 deregulated antisense genes included CTBP1-AS2, OTUD6B-AS1 and AGAP2-AS1, and these findings were replicated in an independent dataset and further validated by qPCR. The expression of these lncRNAs is clearly distinct in patients,

although the functional consequences of these deregulations are at this point difficult to infer given the limited information available on their potential functions. Future in-depth functional analyses are warranted on the functional roles of these genes to confirm their role in SSc pathogenesis. lncRNAs play an important role in development and disease (Batista and Chang 2013; Esteller 2011), but have not yet been described in relation to SSc.

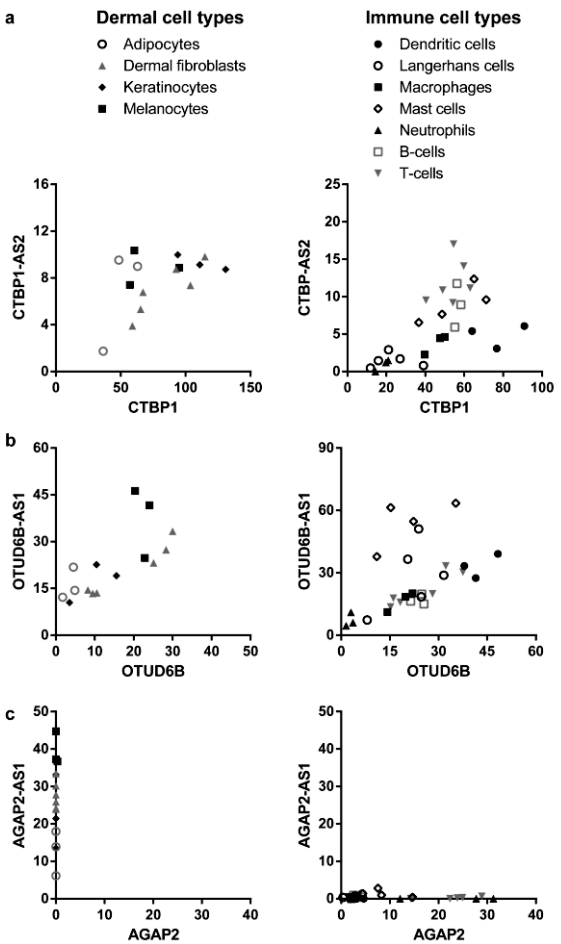

Figure 5. Cell type specific expression of SAS gene pairs in dermal and immune cells. Expression levels for CTBP1 and CTBP1-AS2 (a), OTUD6B and OTUD6B-AS1 (b) and AGAP2 and AGAP2-AS1 (c) in dermal and immune cell types. Expression values are shown as TPM for both the sense and antisense gene. Expression values of each cell type was measured in at least 3 donors. Correlation analysis was performed by spearman rank test.

Most lncRNAs are not yet available on microarrays and are therefore missed in the available data sets that were investigating SSc deregulated genes. More importantly, association of lncRNAs with inflammatory diseases like rheumatoid arthritis, diabetes and psoriasis are increasingly being reported, highlighting their potential role in disease mechanisms (Gupta *et al*. 2016; Messemaker, Huizinga, and Kurreeman 2015). Here, we identify 676 lncRNAs that are deregulated in skin from SSc patients as compared to healthy individuals. A large proportion of the deregulated lncRNAs belonged to the antisense RNA category. Antisense RNAs which reside in a locus with a sense gene (and often span part of this gene) and potentially function as co-regulators of the sense gene (Chan *et al*. 2015; Kimura *et al*. 2013; Peng *et al*. 2015). We identified 26 SAS gene pairs which displayed evidence of differential expression in SSc patients versus controls. From these gene pairs, 55% of the antisense genes showed similar direction of association in an independent data set. The top three deregulated antisense genes included CTBP1-AS2, OTUD6B-AS1 and AGAP2-AS1. OTUD6B is a deubiquitinating enzyme of which little is known. Its downregulation has been linked to cell proliferation in B cells following prolonged cytokine stimulation (Xu *et al*. 2011). CTBP1 is a C terminal binding protein which acts as a transcriptional corepressor and plays a role in epidermal development (Boxer *et al*. 2014). Increased CTBP1 levels were shown to disrupt skin homeostasis (Deng *et al*. 2014). AGAP2 was found upregulated in various cancers and is involved in focal adhesion and cell migration (Jia *et al*. 2016; Zhu *et al*. 2009). Interestingly, AGAP2-AS1 was also shown to be involved in cell migration and is able to repress transcription via interaction with EZH2 and LSD1 in cancer cells (Li *et al*. 2016).

Based on our data, we believe that future studies on functional roles of lncRNAs in SSc pathogenesis might focus on CTBP1-AS2, OTUD6B-AS1 and AGAP2-AS1 as these were significantly deregulated, the deregulation was also found in an independent dataset, and based on current knowledge a role in pathophysiology is plausible. Thereby, one should take into account that we have investigated deregulated polyA-positive lncRNAs, while also polyA-negative lncRNAs exists (Derrien *et al*. 2012). Although polyA-negative lncRNAs are less well-studied, we do hypothesize that also these lncRNAs might play important roles in SSc development and require further investigation (Yang *et al*. 2011). With respect to coding genes, we observe an inflammatory signature, in line with previous research that shows the presence of an interferon/inflammatory signature in early SSc patients (Johnson *et al*. 2015). In contrast to previous research, a clear

TGF-β signal did not come forward from our gene list, despite the increase of fibrosis related-genes as ACTA1 and COMP (Farina *et al*. 2009). When comparing genes from our study with a previous published dataset, a small proportion of genes (n = 619) overlaps suggesting that consistent deregulated genes exist despite SSc-well known disease heterogeneity, large differences in the mean age and disease duration of patients between both studies (Supplementary Table 2). Moreover, an additional comparison with 415 genes obtained from a meta-analysis performed by Lofgren *et al* show that 159 genes overlapped (38%)(Lofgren *et al*. 2016).

We investigated specific SSc-gene sets in more detail to identify genes deregulated in early SSc patients. Our study reports several coding genes which have not previously been highlighted in gene expression studies of SSc. COL4 (COL4A1, COL4A2 and COL4A4), is a gene in the collagen family and is a major component of the dermal-epidermal junction. Elevated levels of COL4 protein have been found in the serum of SSc patients (Gerstmeier H, Gabrielli A, Meurer M, Brocks D, Braun-Falco O 1988) and COL4 autoantibodies have been found in 31% of SSc patients highlighting that an increase of COL4 might play a role in SSc (Riente *et al*. 1995). KRT4 and KRT9, overexpressed genes from our study are normally not expressed in forearm skin. KRT4 is expressed in mucosal tissue and is increased upon inflammation (Bosch *et al*. 1989), while KRT9 is normally expressed in soles and hand palms (Rinn *et al*. 2008). KRT9 is required for structural integrity of the epidermis and KRT9 was found increased in psoriasis patients (Fu *et al*. 2014; Kim *et al*. 2016). The increased expression of these keratins in skin of early SSc patients highlights the possibility of aberrant activation of these genes early in disease.

Besides collagen and keratin genes, we also identified inflammatory genes. Some of these deregulated inflammatory genes are located in loci that are genetically associated to SSc including *HLA* and *TNFAIP3* (Dieudé *et al*. 2010). Interestingly, the expression of TNFAIP3 is strongly reduced in SSc skin tissue. Given the role of TNFAIP3 as a negative regulator of NF-κB signalling, its downregulation would be suggestive of an increased NF-κB activation, possibly further enhancing the increased pro-inflammatory environment. TNFAIP3 was also found deregulated in several other cell types and suggests that genes and pathways are deregulated across multiple tissues(Avouac *et al*. 2011). In line with this, we have also observed clusters of macrophages in our SSc skin biopsies. Also increased CX3CR1 expression came forward and likely contributes to skin

inflammation in SSc as CX3CR1 knockout experiments resulted in decreased skin inflammation (Morimura *et al*. 2016). Interestingly, the top deregulated chemokines were CXCL2 and PF4 (CXCL4). CXCL2, a neutrophil chemoattractant and pro-angiogenic factor (Raman, Sobolik-Delmaire, and Richmond 2011), was reduced and might influence vascular repair within skin of SSc patients (Hummers *et al*. 2009). PF4 (CXCL4) was increased at the RNA level and increased PF4 protein levels were found in SSc serum and skin (van Bon *et al*. 2014). Our study suggests that despite the short disease duration of the patients included in this study, distinct gene expression profiles already exist at an earlier stage in the disease process than investigated so far. Further studies in larger sample sets and long-term follow-up of patients should yield deeper insight into which relevant mechanisms are deregulated in what stage of the disease.

In conclusion, we here report a gene list of 619 genes consistently deregulated over two studies accounting for direction of association and providing a basis of consistent gene expression changes. We show that the expression of keratin genes is increased and that patients display enhanced levels of genes originating from inflammatory gene signatures. In addition, we here provide a blueprint of DE lncRNAs which may play a role as underlying regulators disturbing processes contributing to SSc. Interestingly, even though many of these DE lncRNAs have to our knowledge not yet been described in context of SSc, we show strong correlations with coding genes for several antisense genes. Given the replication in an independent cohort, future studies on the functional role of these specific lncRNAs in SSc pathogenesis are warranted.

MATERIALS AND METHODS

For full details of methods see online supplementary material.

Patient information

Early SSc patients (with a disease duration < 2 years) were recruited at the Department of Rheumatology of the Leiden University Medical Center (Leiden, The Netherlands) and all patients met the American Rheumatism Association classification criteria for SSc (Subcommittee for scleroderma criteria 1980). Patient characteristics can be found in Supplementary Table 3. Institutional review board approval and written informed consent was obtained before patients entered this study. Two 4 mm skin biopsies were taken and from 10

patients the skin biopsy came from a clinically affected area and in 4 patients the skin was locally unaffected. Skin biopsies from healthy individuals were commercially sourced (Tissue Solutions, UK), came from surgeries of arm and leg and were age and sex-matched.

Transcriptome characterisation and analysis

RNA was isolated from skin biopsies and sequenced using polyA selection and a stranded protocol using Ion Torrent next generation sequencing technology (Service XS, The Netherlands). Reads were aligned to the human genome (*Homo sapiens GRh38.78)* using Bowtie2 and STAR and differential expression analysis was carried out using HTseq and DEseq2. All genes with a minimum base mean expression value of 2.3 were included in the differential expression analysis. RNA sequencing files are deposited at the EGA-database under nr: EGAO00000000316 (https://www.ebi.ac.uk/ega/organisations/EGAO00000000316).

Publicly available gene expression datasets and gene sets analysis

DE genes [FDR < 0.05, FC > 1.5] were compared with a publicly available dataset obtained from Assassi *et al*. (Assassi *et al*. 2015). DE genes were investigated via Gene Ontology (GO)-term analysis using Toppgene [version 23 may 2016] and in specific gene sets. Gene sets were obtained from Hugo Gene Nomenclature Committee (HGNC) or by additionally compiled SSc gene sets from alternative sources. Genes in the interferon and alternative macrophage activation signature were obtained from Mahoney *et al*. (Mahoney *et al*. 2015). Genes involved in TGFβ signalling were obtained from the Broad Institute. All genes included in these gene sets are outlined in Supplementary File 1.

Long non-coding RNAs

Genes annotated as lncRNAs (and sub classifications) were obtained from GENCODE (Ensemble version 82) (Harrow *et al*. 2012). Antisense genes were linked to sense genes on the basis of annotations from GENCODE (Harrow *et al*. 2012). Antisense genes with a concomitant DE sense gene were investigated in an as yet unpublished RNA sequencing dataset of skin biopsies of 14 SSc patients and 6 healthy individuals. An overall combined *p*-value was calculated using Fisher's exact test. The top three sense and antisense genes were visualised in IGV to ensure strand specificity and non-overlapping reads (Supplementary Figure 7). The coding potential of antisense genes was determined using an in-

silico coding potential calculator (Kong *et al*. 2007) and analysis of cell specific expression was performed using publicly available FANTOM5 datasets (http://fantom.gsc.riken.jp/5/) (Lizio *et al*. 2015; Severin *et al*. 2014). Correlations between antisense and sense genes were calculated using variance stabilised transformed (VST) counts by spearman rank test.

SUPPLEMENTARY INFORMATION

Supplementary information is available online on the website of the journal of investigative dermatology: Supplementary Figure 1-7 and Supplementary Table 1-4.

ACKNOWLEDGEMENTS

We are grateful for SSc patients for participating in this study. We are also thankful for Koen Quint who aided in the analysis of skin sections. This work was supported by the Dutch Arthritis Foundation, The Netherlands. F. Kurreeman has support from UNESCO-L'Oreal for Women in Science Fellowship, Marie Curie FP7 Outgoing Fellowship and an LUMC Research Fellowship. O. Distler has support by the Swiss National foundation (project number 310030_166259) for studying the role of lncRNAs in SSc.

REFERENCES

Assassi, Shervin, William R Swindell, Minghua Wu, Filemon D Tan, Dinesh Khanna, Daniel E Furst, Donald P Tashkin, *et al*. 2015. "Dissecting the Heterogeneity of Skin Gene Expression Patterns in Systemic Sclerosis." *Arthritis & Rheumatology (Hoboken, N.J.)* 67 (11): 3016–2
6. doi:10.1002/art.39289.

Avouac, Jérôme, Nicolas Cagnard, Jörg H Distler, Yoland Schoindre, Barbara Ruiz, Pierre Olivier Couraud, Georges Uzan, Catherine Boileau, Gilles Chiocchia, and Yannick Allanore. 2011. "Insights into the Pathogenesis of Systemic Sclerosis Based on the Gene Expression Profile of Progenitor-Derived Endothelial Cells." *Arthritis and Rheumatism* 63 (11): 3552–62. doi:10.1002/art.30536.

Batista, Pedro J, and Howard Y Chang. 2013. "Long Noncoding RNAs: Cellular Address Codes in Development and Disease." *Cell* 152 (6). Elsevier Inc.: 1298–1307. doi:10.1016/j.cell.2013.02.012.

Bon, Lenny van, Alsya J Affandi, Jasper Broen, Romy B Christmann, Renoud J Marijnissen, Lukasz Stawski, Giuseppina a Farina, *et al*. 2014. "Proteome-Wide Analysis and CXCL4 as a Biomarker in Systemic Sclerosis." *The New England Journal of Medicine* 370 (5): 433–43. doi:10.1056/NEJMoa1114576.

Bosch, Franz X, Jean-pierre Ouhayoun, Bernhard L Bader, Christine Collin, Christine Grund, Inchul Lee, and Werner W Franke. 1989. "Extensive Changes in Cytokeratin Expression Patterns in Pathologically Affected Human Gingiva." *Virchows Archiv B Cell Pathology*, 59–77.

Boxer, Lisa D, Brook Barajas, Shiying Tao, Jiajing Zhang, and Paul A Khavari. 2014. "ZNF750 Interacts with KLF4 and RCOR1 , Regulators to Repress Epidermal Progenitor Genes and Induce Differentiation Genes," 2013–26. doi:10.1101/gad.246579.114.Bao.

Chan, Jennie, Maninjay Atianand, Zhaozhao Jiang, Susan Carpenter, Daniel Aiello, Roland Elling, Katherine a Fitzgerald, and Daniel R Caffrey. 2015. “Cutting Edge: A Natural Antisense Transcript, AS-IL1α, Controls Inducible Transcription of the Proinflammatory Cytokine IL-1α.” *Journal of Immunology (Baltimore, Md. : 1950)* 195 (4): 1359–63. doi:10.4049/jimmunol.1500264.

Deng, Hui, Fulun Li, Hong Li, Yu Deng, Jing Liu, Donna Wang, Gangwen Han, Xiao-Jing Wang, and Qinghong Zhang. 2014. “CtBP1 Overexpression in Keratinocytes Perturbs Skin Homeostasis.” *The Journal of Investigative Dermatology* 134 (5). Elsevier Masson SAS: 1323–31. doi:10.1038/jid.2013.504.

Derrien, Thomas, Rory Johnson, Giovanni Bussotti, Andrea Tanzer, Sarah Djebali, Hagen Tilgner, Gregory Guernec, *et al*. 2012. “The GENCODE v7 Catalog of Human Long Noncoding RNAs : Analysis of Their Gene Structure , Evolution , and Expression.” *Genome Research*, 1775–89. doi:10.1101/gr.132159.111.

Dieudé, P, M Guedj, J Wipff, B Ruiz, G Riemekasten, M Matucci-Cerinic, I Melchers, *et al*. 2010. “Association of the TNFAIP3 rs5029939 Variant with Systemic Sclerosis in the European Caucasian Population.” *Annals of the Rheumatic Diseases* 69 (11): 1958–64. doi:10.1136/ard.2009.127928.

Esteller, Manel. 2011. “Non-Coding RNAs in Human Disease.” *Nature Reviews. Genetics* 12 (12). Nature Publishing Group: 861–74. doi:10.1038/nrg3074.

Faghihi, Mohammad Ali, and Claes Wahlestedt. 2010. “Regulatory Roles of Natural Antisense Transcripts” 10 (9): 637–43. doi:10.1038/nrm2738.Regulatory.

Farina, G, R Lemaire, P Pancari, J Bayle, R L Widom, and R Lafyatis. 2009. “Cartilage Oligomeric Matrix Protein Expression in Systemic Sclerosis Reveals Heterogeneity of Dermal Fibroblast Responses to Transforming Growth Factor Beta.” *Annals of the Rheumatic Diseases* 68 (3): 435–41. doi:10.1136/ard.2007.086850.

Fu, Dun Jack, Calum Thomson, Declan P Lunny, Patricia J Dopping-Hepenstal, John a McGrath, Frances J D Smith, W H Irwin McLean, and Deena M Leslie Pedrioli. 2014. “Keratin 9 Is Required for the Structural Integrity and Terminal Differentiation of the Palmoplantar Epidermis.” *The Journal of Investigative Dermatology* 134 (3). Elsevier Masson SAS: 754–63. doi:10.1038/jid.2013.356.

Gardner, Humphrey, Jeffrey R Shearstone, Raj Bandaru, Tom Crowell, Matthew Lynes, Maria Trojanowska, Jaspreet Pannu, *et al*. 2006. “Gene Profiling of Scleroderma Skin Reveals Robust Signatures of Disease That Are Imperfectly Reflected in the Transcript Profiles of Explanted Fibroblasts.” *Arthritis and Rheumatism* 54 (6): 1961–73. doi:10.1002/art.21894.

Gerstmeier H, Gabrielli A, Meurer M, Brocks D, Braun-Falco O, Krieg T. 1988. “Levels of Type IV Collagen and Laminin Fragments in Serum from Patients with Progressive Systemic Sclerosis.” *J Rheumatol.* Jun;15(6):

Greenblatt, Matthew B, Jennifer L Sargent, Giuseppina Farina, Kelly Tsang, Robert Lafyatis, Laurie H Glimcher, Michael L Whitfield, and Antonios O Aliprantis. 2012. “Interspecies Comparison of Human and Murine Scleroderma Reveals IL-13 and CCL2 as Disease Subset-Specific Targets.” *The American Journal of Pathology* 180 (3). Elsevier Inc.: 1080–94. doi:10.1016/j.ajpath.2011.11.024.

Gupta, Rashmi, Richard Ahn, Kevin Lai, Elizabeth Mullins, Maya Debbaneh, Michelle Dimon, Sarah Arron, and Wilson Liao. 2016. “Landscape of Long Noncoding RNAs in Psoriatic and Healthy Skin.” *The Journal of Investigative Dermatology* 136 (3). The Authors: 603–9. doi:10.1016/j.jid.2015.12.009.

Harrow, Jennifer, Adam Frankish, Jose M Gonzalez, Electra Tapanari, Mark Diekhans, Felix Kokocinski, Bronwen L Aken, *et al*. 2012. “GENCODE: The Reference Human Genome Annotation for the ENCODE Project.” *Genome Research* 22 (9): 1760–74. doi:10.1101/gr.135350.111.

Hummers, Laura K, Amy Hall, Fredrick M Wigley, and Michael Simons. 2009. “Abnormalities in the Regulators of Angiogenesis in Patients with Scleroderma.” *The Journal of Rheumatology* 36 (3): 576–82. doi:10.3899/jrheum.080516.

Jia, Weijuan, Y I Feng, Andrew J Sanders, Eleri L Davies, and W E N G Jiang. 2016. “Phosphoinositide-3-Kinase Enhancers , PIKEs : Their Biological Functions and Roles in Cancer” 1110: 1103–9.

Johnson, Michael E., J. Matthew Mahoney, Jaclyn Taroni, Jennifer L. Sargent, Eleni Marmarelis, Ming-Ru Wu, John Varga, Monique E. Hinchcliff, and Michael L. Whitfield. 2015. “Experimentally-Derived Fibroblast Gene Signatures Identify Molecular Pathways Associated with Distinct Subsets of Systemic Sclerosis Patients in Three

Independent Cohorts." Edited by Stamatis-Nick Liossis. *Plos One* 10 (1): e0114017. doi:10.1371/journal.pone.0114017.

Katayama, S, Y Tomaru, T Kasukawa, K Waki, M Nakanishi, M Nakamura, H Nishida, *et al*. 2005. "Antisense Transcription in the Mammalian Transcriptome." *Science (New York, NY)* 309 (5740): 1564–66. doi:10.1126/science.1112009.

Kim, Dongwon, M Zulfiquer Hossain, Ashley Nieves, Lihong Gu, Tabetha S Ratliff, Seung Mi Oh, Angela Park, *et al*. 2016. "To Control Site-Specific Skin Gene Expression, Autocrine Mimics Paracrine Canonical Wnt Signaling and Is Activated Ectopically in Skin Disease." *The American Journal of Pathology* 186 (5): 1140–50. doi:10.1016/j.ajpath.2015.12.030.

Kimura, Tominori, Shiwen Jiang, Mikio Nishizawa, and Emi Yoshigai. 2013. "Stabilization of Human Interferon- a 1 mRNA by Its Antisense RNA," 1451–67. doi:10.1007/s00018-012-1216-x.

Kong, Lei, Yong Zhang, Zhi-Qiang Ye, Xiao-Qiao Liu, Shu-Qi Zhao, Liping Wei, and Ge Gao. 2007. "CPC: Assess the Protein-Coding Potential of Transcripts Using Sequence Features and Support Vector Machine." *Nucleic Acids Research* 35 (Web Server issue): W345-9. doi:10.1093/nar/gkm391.

Li, W, M Sun, C Zang, P Ma, J He, M Zhang, Z Huang, Y Ding, and Y Shu. 2016. "Upregulated Long Non-Coding RNA AGAP2-AS1 Represses LATS2 and KLF2 Expression through Interacting with EZH2 and LSD1 in Non-Small-Cell Lung Cancer Cells." *Cell Death & Disease* 7 (300): e2225. doi:10.1038/cddis.2016.126.

Lizio, Marina, Jayson Harshbarger, Hisashi Shimoji, Jessica Severin, Takeya Kasukawa, Serkan Sahin, Imad Abugessaisa, *et al*. 2015. "Gateways to the FANTOM5 Promoter Level Mammalian Expression Atlas." *Genome Biology* 16 (January): 22. doi:10.1186/s13059-014-0560-6.

Lofgren, Shane, Monique Hinchcliff, Mary Carns, Tammara Wood, Kathleen Aren, Esperanza Arroyo, Peggie Cheung, *et al*. 2016. "Integrated, Multicohort Analysis of Systemic Sclerosis Identifies Robust Transcriptional Signature of Disease Severity." *JCI Insight* 1 (21): e89073. doi:10.1172/jci.insight.89073.

Mahoney, J Matthew, Jaclyn Taroni, Viktor Martyanov, Tammara a Wood, Casey S Greene, Patricia a Pioli, Monique E Hinchcliff, and Michael L Whitfield. 2015. "Systems Level Analysis of Systemic Sclerosis Shows a Network of Immune and Profibrotic Pathways Connected with Genetic Polymorphisms." *PLoS Computational Biology* 11 (1): e1004005. doi:10.1371/journal.pcbi.1004005.

Mathes, Allison L, Romy B Christmann, Giuseppina Stifano, Alsya J Affandi, Timothy R D J Radstake, G Alessandra Farina, Cristina Padilla, Sarah McLaughlin, and Robert Lafyatis. 2014. "Global Chemokine Expression in Systemic Sclerosis (SSc): CCL19 Expression Correlates with Vascular Inflammation in SSc Skin." *Annals of the Rheumatic Diseases* 73 (10): 1864–72. doi:10.1136/annrheumdis-2012-202814.

Melé, Marta, and John L Rinn. 2016. "'Cat's Cradling' the 3D Genome by the Act of LncRNA Transcription." *Molecular Cell* 62 (5): 657–64. doi:10.1016/j.molcel.2016.05.011.

Messemaker, Tobias C, Tom W Huizinga, and Fina Kurreeman. 2015. "Immunogenetics of Rheumatoid Arthritis: Understanding Functional Implications." *Journal of Autoimmunity* 64 (July). Elsevier Ltd: 7–14. doi:10.1016/j.jaut.2015.07.007.

Milano, Ausra, Sarah a Pendergrass, Jennifer L Sargent, Lacy K George, Timothy H McCalmont, M Kari Connolly, and Michael L Whitfield. 2008. "Molecular Subsets in the Gene Expression Signatures of Scleroderma Skin." *PloS One* 3 (7): e2696. doi:10.1371/journal.pone.0002696.

Morimura, Sohshi, Tomonori Oka, Makoto Sugaya, and Shinichi Sato. 2016. "CX3CR1 Deficiency Attenuates Imiquimod-Induced Psoriasis-like Skin Inflammation with Decreased M1 Macrophages." *Journal of Dermatological Science* 82 (3). Japanese Society for Investigative Dermatology: 175–88. doi:10.1016/j.jdermsci.2016.03.004.

Pelechano, Vicent, and Lars M Steinmetz. 2013. "Gene Regulation by Antisense Transcription." *Nature Reviews. Genetics* 14 (12). Nature Publishing Group: 880–93. doi:10.1038/nrg3594.

Pendergrass, Sarah a, Raphael Lemaire, Ian P Francis, J Matthew Mahoney, Robert Lafyatis, and Michael L Whitfield. 2012. "Intrinsic Gene Expression Subsets of Diffuse Cutaneous Systemic Sclerosis Are Stable in Serial Skin Biopsies." *The Journal of Investigative Dermatology* 132 (5). Elsevier Masson SAS: 1363–73. doi:10.1038/jid.2011.472.

Peng, Huiyong, Yingzhao Liu, Jie Tian, Jie Ma, Xinyi Tang, Ke Rui, and Xinyu Tian. 2015. “The Long Noncoding RNA IFNG-AS1 Promotes T Helper Type 1 Cells Response in Patients with Hashimoto ’ S Thyroiditis.” *Nature Publishing Group*, no. December. Nature Publishing Group: 1–9. doi:10.1038/srep17702.

Raman, Dayanidhi, Tammy Sobolik-Delmaire, and Ann Richmond. 2011. “Chemokines in Health and Disease.” *Experimental Cell Research* 317 (5). Elsevier B.V.: 575–89. doi:10.1016/j.yexcr.2011.01.005.

Rau, Andrea, Guillemette Marot, and Florence Jaffrézic. 2014. “Differential Meta-Analysis of RNA-Seq Data from Multiple Studies,” 1–10.

Richards, Edward J, Gu Zhang, Zhu-peng Li, Jennifer Permuth-wey, Sridevi Challa, Yajuan Li, William Kong, *et al*. 2015. “Long Non-Coding RNAs (LncRNA) Regulated by Transforming Growth Factor (TGF) ” 290 (11): 6857–67. doi:10.1074/jbc.M114.610915.

Riente, L, B Marchini, M P Dolcher, A Puccetti, S Bombardieri, and P Migliorini. 1995. “Anti-Collagen Antibodies in Systemic Sclerosis and in Primary Raynaud â€™ S Phenomenon.” *Clin. Exp. Immunol.*, 354–59.

Rinn, John L, Jordon K Wang, Helen Liu, Kelli Montgomery, Matt van de Rijn, and Howard Y Chang. 2008. “A Systems Biology Approach to Anatomic Diversity of Skin.” *The Journal of Investigative Dermatology* 128 (4). Elsevier Masson SAS: 776–82. doi:10.1038/sj.jid.5700986.

Rinn JL, Chang HY. 2013. “Genome Regulation by Long Noncoding RNAs.” *Annu. Rev. Biochem.*, 8–12. doi:10.1146/annurev-biochem-051410-092902.Genome.

Rubio-Rivas, Manuel, Cristina Royo, Carmen Pilar Simeón, Xavier Corbella, and Vicent Fonollosa. 2014. “Mortality and Survival in Systemic Sclerosis: Systematic Review and Meta-Analysis.” *Seminars in Arthritis and Rheumatism* 44 (2). Elsevier: 208–19. doi:10.1016/j.semarthrit.2014.05.010.

Severin, Jessica, Marina Lizio, Jayson Harshbarger, Hideya Kawaji, Carsten O Daub, Yoshihide Hayashizaki, Nicolas Bertin, and Alistair R R Forrest. 2014. “Interactive Visualization and Analysis of Large-Scale Sequencing Datasets Using ZENBU.” *Nature Biotechnology* 32 (3): 217–19. doi:10.1038/nbt.2840.

Subcommittee for scleroderma criteria, 1980. 1980. “Preliminary Criteria for the Classification of Systemic Sclerosis (Scleroderma).” *Arthritis and Rheumatism* 23 (5): 581–90. doi:10.1002/art.1780230510.

Villegas, Victoria E, and Peter G Zaphiropoulos. 2015. “Neighboring Gene Regulation by Antisense Long Non-Coding RNAs.” *International Journal of Molecular Sciences* 16 (2): 3251–66. doi:10.3390/ijms16023251.

Werner, Andreas. 2013. “Biological Functions of Natural Antisense Transcripts.” *BMC Biology*, 2–4.

Whitfield, Michael L, Deborah R Finlay, John Isaac Murray, Olga G Troyanskaya, Jen-tsan Chi, Alexander Pergamenschikov, Timothy H McCalmont, Patrick O Brown, David Botstein, and M Kari Connolly. 2003. “Systemic and Cell Type-Specific Gene Expression Patterns in Scleroderma Skin.” *Proceedings of the National Academy of Sciences of the United States of America* 100 (21): 12319–24. doi:10.1073/pnas.1635114100.

Xu, Zhongping, Yufang Zheng, Yufei Zhu, Xiangyin Kong, and Landian Hu. 2011. “Evidence for OTUD-6B Participation in B Lymphocytes Cell Cycle after Cytokine Stimulation.” *PloS One* 6 (1): e14514. doi:10.1371/journal.pone.0014514.

Yang, Li, Michael O Duff, Brenton R Graveley, Gordon G Carmichael, and Ling-ling Chen. 2011. “Genomewide Characterization of Non-Polyadenylated RNAs,” 1–14.

Zhu, Yunjuan, Yuanjun Wu, Jae I Kim, Zhimin Wang, Yehia Daaka, and Zhongzhen Nie. 2009. “Arf GTPase-Activating Protein AGAP2 Regulates Focal Adhesion Kinase Activity and Focal Adhesion Remodeling.” *The Journal of Biological Chemistry* 284 (20): 13489–96. doi:10.1074/jbc.M900469200.

Chapter 7

Allele-specific repression of *Sox2* through the long non-coding RNA *Sox2ot*

Sci Rep. 2018 Jan 10;8(1):386. doi: 10.1038/s41598-017-18649-4.

Tobias C. Messemaker, Selina M. van Leeuwen, Patrick van den Berg, Anke J. ‘t Jong, Robert-Jan Palstra, Rob C. Hoeben, Stefan Semrau and Harald M.M. Mikkers

Abstract

The transcription factor Sox2 controls the fate of pluripotent stem cells and neural stem cells. This gatekeeper function requires well-regulated Sox2 levels. We postulated that *Sox2* regulation is partially controlled by the *Sox2* overlapping long non-coding RNA (lncRNA) gene *Sox2ot*. Here we show that the RNA levels of *Sox2ot* and *Sox2* are inversely correlated during neural differentiation of mouse embryonic stem cells (ESCs). Through allele-specific enhanced transcription of *Sox2ot* in mouse *Sox2eGFP* knockin ESCs we demonstrate that increased *Sox2ot* transcriptional activity reduces *Sox2* RNA levels in an allele-specific manner. Enhanced *Sox2ot* transcription, yielding lower *Sox2* RNA levels, correlates with a decreased chromatin interaction of the upstream regulatory sequence of *Sox2* and the ESC-specific *Sox2* super enhancer. Our study indicates that, in addition to previously reported in trans mechanisms, *Sox2ot* can regulate *Sox2* by an allele-specific mechanism, in particular during development.

Introduction

Correct gene regulation, which relies on the temporally and spatially controlled expression of lineage specific transcription factors, determines the success of development. Sox2 is such a transcription factor key to development. *Sox2* belongs to the family of high mobility group (HMG) DNA binding domain genes related to the sex determining gene Y (Sry) and together with *Sox1* and *Sox3*, *Sox2* forms the SoxB1 family. *Sox2* exerts its cell type specific function by interaction with other homeodomain transcription factors, the POU domain protein Oct4, or the paired domain protein Pax6[1]. An important function of *Sox2* is maintaining the stem cell state of either naïve or primed pluripotent stem cells[2]. Reduction or overexpression of *Sox2* in mouse and human embryonic stem cells (ESCs) induces the differentiation into primarily endoderm and trophoectoderm-like cells, respectively[3–8]. Endogenous *Sox2* levels also influence the germ layer fate of pluripotent stem cells. High endogenous levels steer pluripotent cells into the (neural) ectodermal lineage, whereas low levels promote mesendodermal differentiation[9]. *Sox2* fulfills a similar role in neural stem cells (NSCs) *in vitro* and *in vivo*. Overexpression of *Sox2* in NSCs of the developing spinal cord represses differentiation by counteracting transcription factor driven proneural programs, whereas *Sox2* protein inhibition enhances differentiation[10,11]. In the developing eye, retinal progenitor cells lose their

proliferation and differentiation capacity after *Sox2* ablation[12]. Reduced *Sox2* levels (<40%) cause microphthalmia due to aberrant differentiation of the progenitor cells[12]. In addition, misexpression of *Sox2* in astrocytes converts them into neuroblasts[13], whereas it activates neural transcription programs in cells of mesodermal origin[14,15]. Thus, well-controlled and tightly-timed *Sox2* activity appears to be important for correct neural development.

Sox2 activity is controlled by post-translational modifications, such as serine- and threonine phosphorylation, sumoylation, ubiquitination, and acytelation[16]. These modifications affect localization, DNA binding and stability. However, *Sox2* activity is to a great extent controlled at the transcriptional level. The requirement for well-balanced, tightly controlled, and cell type specific expression explains the complex genomic architecture of the *Sox2* locus. Multiple enhancer elements that drive tissue specific expression have been identified in the 200 kb region surrounding *Sox2*[17–20]. Consequently, endogenous expression has only been fully recapitulated in transgenic mice through a knockin approach where one of the *Sox2* alleles was replaced by a marker gene[12,21,22] or through introduction of bacterial artificial chromosomes (BACs) containing >200 kb of *Sox2* genomic sequences[23].

Protein encoding genes like transcription factors and chromatin modifiers are key to transcription activation. However, RNA genes that do not encode proteins can fulfill transcriptional regulatory roles as well. Long non-coding RNAs (lncRNAs), which are >200 nucleotides in length, seem to have in particular evolved for controlling genes at a transcriptional level[24]. LncRNA-mediated transcription regulation is instructed in cis or in trans. Allele-specific in cis mechanisms include recruitment of chromatin modifying complexes repressing transcription[25] or activating transcription[26], transcriptional interference preventing transcription factor access[27,28], or gene looping[29]. Recently, a lncRNA gene called *Sox2* overlapping transcript (*Sox2ot*) that is transcribed in the same direction as *Sox2* and is polyadenylated downstream of *Sox2* was described[30,31]. To date several studies investigating the function of *Sox2ot* have been reported[32–34]. These studies utilized knockdown or overexpression of *Sox2ot* in cancer cell lines and the results have indicated a role of *Sox2ot* in regulating proliferation as well as regulating *Sox2*. *Sox2ot* levels were invariably positively correlated with *Sox2*, however, the underlying regulatory mechanism has remained unknown.

In this study we evaluated expression of *Sox2ot* during development and studied the effect of *Sox2ot* overexpression in modified mouse ESCs that allow discrimination between cis and trans regulatory effects. On basis of our data we propose that during development *Sox2ot* expression is mainly restricted to neural cell types and that, in contrast to previous reports, enhanced *Sox2ot* transcriptional activity negatively affects *Sox2* RNA levels in an allele-specific manner.

Results

Characterization and conservation of Sox2ot transcripts

The *Sox2* gene is a single exon gene that is located in a gene desert on mouse chromsosome 3 (Fig. 1a). Apart from *Sox2* the only genes located within a 200 kb stretch of genomic DNA are presumably of non-coding nature. Two lncRNA genes (*Sox2otb* and *Sox2otc*) have been identified in this region[31]. The transcripts are initiated (~88 kb and ~11 kb) upstream of *Sox2* and are terminated ~40 kb downstream of *Sox2* (Fig. 1a). Transcriptome data, such as ESTs (expressed sequence tags) representing either *Sox2ot* transcript, have indicated that *Sox2ot* transcripts, like the flanking *Sox2* gene, are predominantly present in brain as well as cell lines of neural origin. The expression pattern points to a function of *Sox2ot* in neural development and neural physiology, possibly through a *Sox2*-related mechanism. We first validated the transcription *Sox2ot* genes in neural progenitor cells (NPCs) derived from the lateral wall of the lateral ventricle in adult mouse. Primers recognizing an exon of *Sox2otb* that also is the first exon of *Sox2otc* could amplify *Sox2ot* transcripts in early passage neurospheres (data not shown and Fig. 1g), which is in agreement with two recent studies[31,35]. Using 5′ RLM-RACE we confirmed the 5′ ends of *Sox2otb* and *Sox2otc* (Supplementary Fig. S1a). Full-length cDNA sequence analysis showed extensive splicing, which is arguably random as almost any possible exon conjunction was retrieved. The splicing is largely conserved in other mammals as was recently shown[34]. We identified one previously undescribed exon located between *Sox2otb* exon 2 and *Sox2otc* exon 1 (Fig. 1a). We analyzed the cDNA sequences for the presence of open reading frames (ORFs) through Coding Potential Calculator[36], NCBI's ORFfinder, and a translation initiation prediction program (ATGpr) but the outcome underscored the non-coding nature of all *Sox2otb* and *Sox2otc* splice variants (Supplementary Fig. S1d,e, and f). To test whether the transcripts can be

translated into a polypeptide we performed in vitro transcription/translation assays using the largest, multi-exonic, *Sox2otb* and *Sox2otc* cDNA sequences, but we could not detect any *Sox2*ot polypeptides (Supplementary Fig. S1g). This result indicates that *Sox2otb* and *Sox2otc* are likely of non-coding nature as was suggested before[30,31]. However, our analyses do not fully exclude the generation of very small peptides with a function, which can be produced from presumed non-coding RNA transcripts[37].

Sox2ot exonic and intronic sequences have been conserved between mammals and vertebrates (Supplementary Fig. S1c)[31]. The extent of conservation of genomic sequences between man and other vertebrates, like marsupials, is a measure of importance of these sequences for development. A larger evolutionary distance, i.e. between man and pufferfish (*Fugu rubripes*) diverging 450 million years ago, has been shown to be even more instrumental in uncovering coding as well as non-coding sequences crucial for proper development[38]. It was previously reported that the highest level of evolutionary conservation was observed in the promoter proximal regions of lncRNAs[39–41]. Likewise, the regions surrounding *Sox2otb* exon 1 and *Sox2otc* exon 1, and not the exonic sequences, are highly conserved between man and *Fugu*. The high conservation of *Sox2ot* proximal promoter regions infers that *Sox2ot* sequences that govern transcription are more important during development than the transcript per se.

Expression of Sox2ot during neural development

Since previous studies have indicated that *Sox2ot* expression positively correlates with *Sox2* RNA, we wished to test the correlative expression during neural development. We restricted the expression analysis to *Sox2otb*, *Sox2otc* and *Sox2* only. First we analyzed expression of *Sox2otb*, *Sox2otb* and *Sox2otc* (from here on referred to as *Sox2otb/c* because the riboprobe contains *Sox2otc* exon 1 sequence, which is also present *Sox2otb* transcripts), and *Sox2* in developing mouse embryos using RNA whole mount in situ hybridization (ISH). At 9.25 dpc *Sox2* expression is mainly restricted to the neural tube, developing brain, nasal placodes, otic vesicles and optic vesicles (Fig. 1b, Supplementary Fig. S2a,b) (sense controls in Supplementary Fig. S2c). In contrast, a probe recognizing *Sox2otb* showed an expression pattern limited to the ventral part of the neural tube and optic vesicle, whereas a probe hybridizing to *Sox2otb/c* showed additional expression in the developing brain and otic vesicles (Fig. 1b).

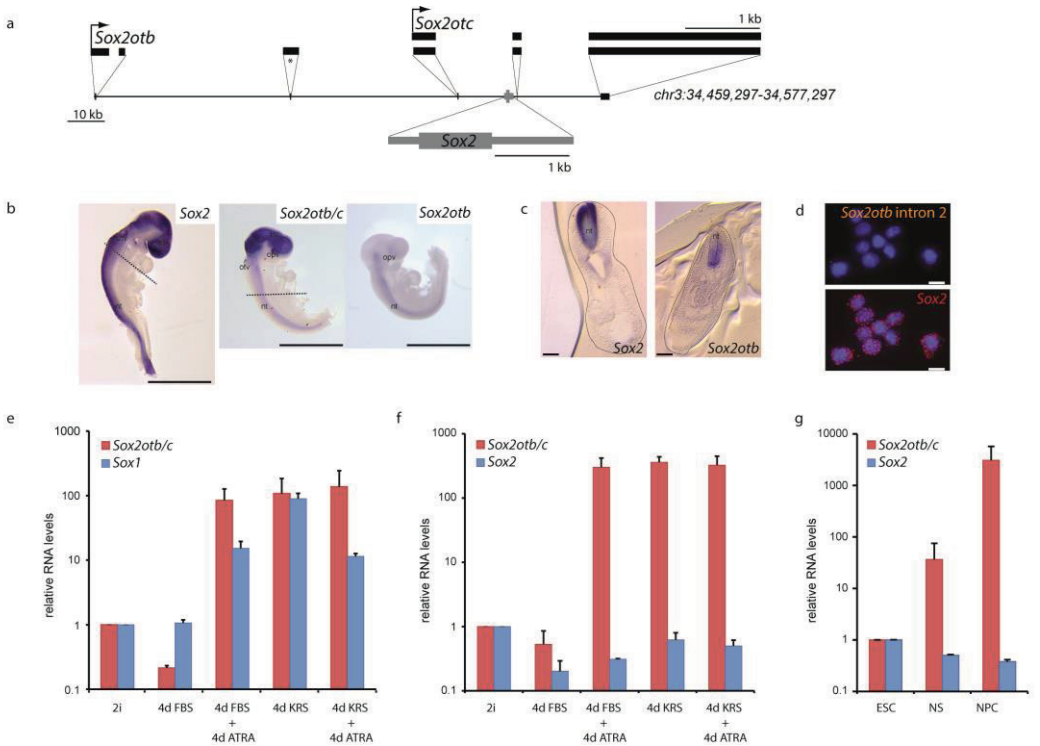

Figure 1. Co-expression of *Sox2otb/c* and *Sox2* during mouse neural development. (a) Schematic representation of the *Sox2* locus on mouse chromosome 3 (mm9 assembly). Depicted are the single exon gene *Sox2*, and the overlapping *Sox2otb* and Soxtotc genes. *Sox2otb* shares exons 4, 5 and 6 with *Sox2otc*. *Indicates a newly identified exon. (b) Whole mount RNA in situ hybridization of E9.25 mouse embryos using antisense *Sox2*, *Sox2otb/c* and *Sox2otb* RNA probes. Scale bar represents 1 mm. otv, otic vesicle; opv, optic vesicle; nt, neural tube; bv, brain vesicle. (c) Transverse sections of the embryos depicted in b. Dashed line in b indicates the level of the transverse section. nt, neural tube. Scale bar represents 100 µm. (d) smFISH on mouse ESCs using *Sox2otb* intron 2 (upper panel) or *Sox2* (lower panel) probe sets. Nuclei are stained with DAPI. (e) and (f) qRT-PCR analysis of *Sox2otb/c* and *Sox1* (e), or *Sox2otb/c* and *Sox2* (f) RNA levels during EB-mediated neural differentiation of mouse ESCs. Cells were cultured for 4 days in FBS or KSR containing medium followed by another 4 days in the same medium with 0.5 µM ATRA. (g) qRT-PCR analysis of *Sox2otb/c* and *Sox2* RNA levels in mouse ESCs, ESC-derived radial glia-like NS cells and NPCs derived from the lateral wall of the lateral ventricle of the adult mouse. Expression was first normalized against β-Actin (e and f) or Myl6 (g), after which the relative expression to the expression in mouse ESCs was calculated. Values are mean + standard deviation (SD) of one representative out of 3 experiments and presented on a 10 log scale.

The spatial and temporal specific expression patterns of *Sox2otb* and *Sox2otc* during neural development indicate that the independent *Sox2ot* transcripts may have different roles. Although it is difficult to robustly interpret co-localization data at the single cell level on basis of RNA ISH using independent single probe

hybridizations, the ISH data show that *Sox2otb*, *Sox2otc* and *Sox2* are co-localized in tissues during neural development.

To further investigate *Sox2otb*, *Sox2otc* and *Sox2* coexpression we analyzed *Sox2otb/c* and *Sox2* expression during the differentiation of mouse ESCs into neuroectoderm. In the tested feeder-independent and feeder-dependent wild type mouse ESC lines *Sox2otb/c* is very lowly expressed during maintenance. This is in sharp contrast with a previous study, which claimed abundant expression of *Sox2ot* in ESCs[31]. To further corroborate the low level of *Sox2ot* expression in ESCs we measured transcription of *Sox2ot* in mouse ESCs by single molecule FISH (smFISH) using a probe set lying in intron 2 of *Sox2otb*. smFISH has single molecule sensitivity[42], yet, *Sox2otb* transcripts were very rare confirming the qRT-PCR results (Fig. 1d, positive control in Supplementary Fig. S2d). We observed a strong upregulation of *Sox2otb/c* upon neurectodermal differentiation using embryoid bodies (Fig. 1e, and Supplementary Fig. S2e). Upregulation coincides with the presence of neural progenitor/stem cells (NP/SCs) as measured through induction of *Sox1*, which is a very early and specific marker of the neuroectoderm lineage[43]. *Sox2ot* induction is all trans retinoic acid (ATRA) independent as neuroectodermal differentiation using knockout replacement serum (KRS) that is devoid of any form of retinol yielded a similar induction of *Sox2otb/c* (Fig. 1e).

In more defined monolayer-based differentiation conditions *Sox2otb/c* was also induced upon neural differentiation (Supplementary Fig. S2d, and f), whereas BMP4-mediated differentiation towards mesendoderm failed to induce *Sox2otb/c* RNA levels (Supplementary Fig. S2g) indicating a primary role of *Sox2ot* in neural development. These results differ from the observations by Amaral *et al*., who have reported higher *Sox2ot* expression levels in mouse ESCs and enhanced *Sox2ot* transcription upon mesodermal commitment[31]. The discrepancies may be caused by differences in the used maintenance and differentiation protocols. Alternatively, a confounding factor may have been transcription initiation downstream of *Sox2* in certain cell types, which yields transcripts that encompass *Sox2ot* exon 6 sequences.

ESC-based neural differentiation cultures are a mixture of distinct cell types, which include ESCs, NSCs/NPCs, and early neurons. During neural differentiation

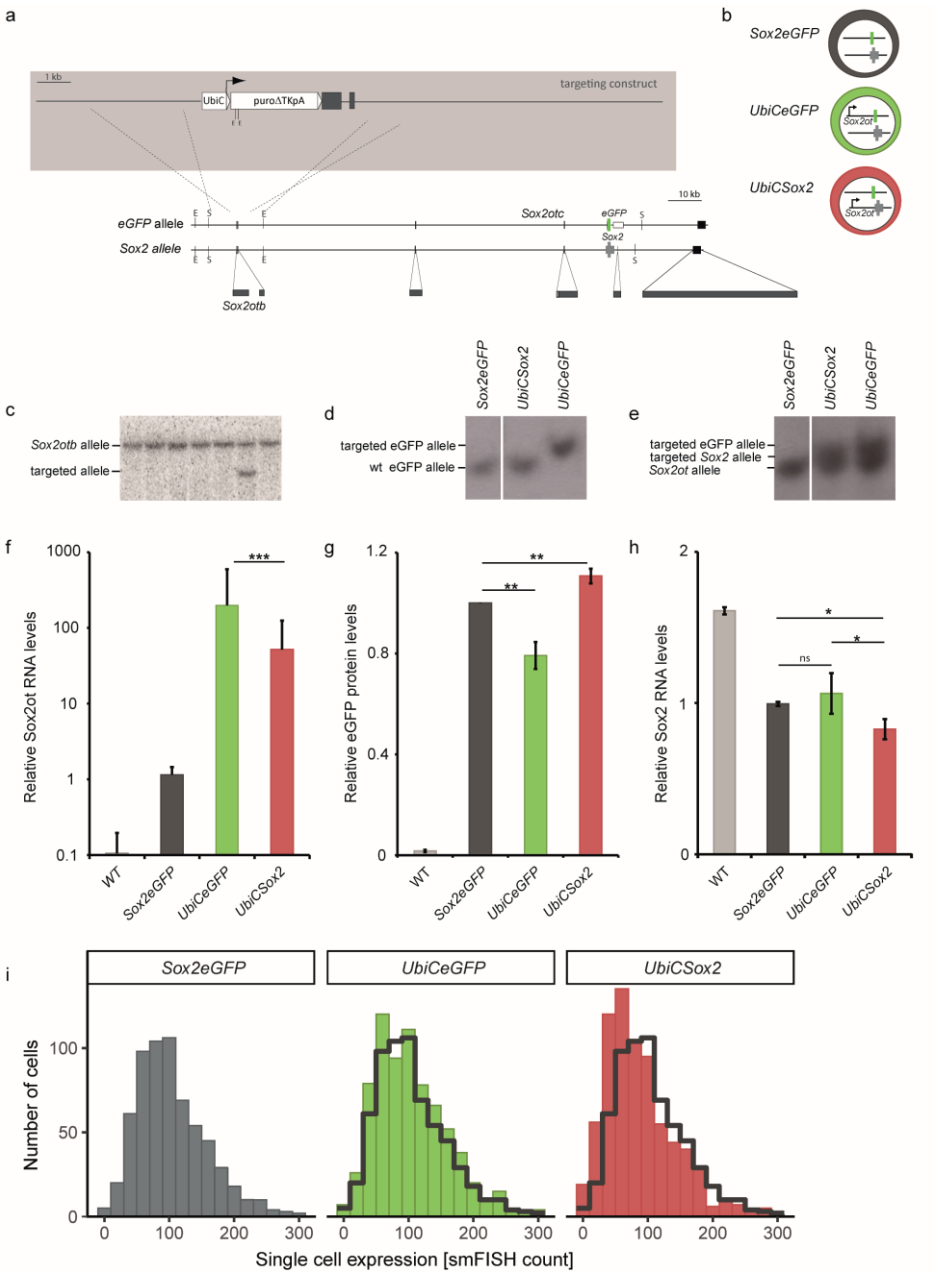

Figure 2. Allele-specific overexpression of endogenous *Sox2otb*. (a) Schematic view of the targeting strategy and targeting construct to generate allele-specific transcription of *Sox2ot*. R = EcoRV and S = SbfI restriction sites. (b) Illustration of the genetic possibilities after targeting the *Sox2eGFP* ESC line: *Sox2eGFP* (untargeted), *UbiCeGFP* (*Sox2ot* is expressed from the *eGFP* allele), or *UbiCSox2* (*Sox2ot* is expressed from the *Sox2* allele). (c) Southern blotting showing correctly recombined *Sox2eGFP* ESC clone using a 3' probe (EcoRV restricted DNA). (d) and (e) Southern blot analysis showing correct targeting of the *eGFP* allele (*UbiCeGFP*) or *Sox2* (*UbiCSox2*) allele using eGFP (d) or

Sox2 (e) specific probes (SbfI restricted DNA). Full blots are shown in Supplementary Fig. S3a. (f) *Sox2otb* expression in *Sox2eGFP*, *UbiCeGFP*, and *UbiCSox2* cells as measured by qRT-PCR. (g) eGFP expression measured by flow cytometry in *Sox2eGFP*, *UbiCeGFP*, and *UbiCSox2* cells. (h) *Sox2* RNA levels in *Sox2eGFP*, *UbiCeGFP*, and *UbiCSox2* cells measured by qRT-PCR. (i) smFISH quantification of *Sox2* RNA copies per single cell in *Sox2eGFP*, *UbiCeGFP*, and *UbiCSox2* lines. The gray line depicts the distribution of *Sox2* in *Sox2eGFP* cells. ***P value < 0.002, **P value < 0.01 *P value < 0.05. Results are from three independent experiments using (sub)clones of *Sox2eGFP* (n = 2), *UbiCeGFP* (n = 3), and *UbiCSox2* (n = 2). Values are presented as mean +/− SD (g and h) or +SD (10 log scale (f)). qRT-PCR data were normalized against β-Actin, and relative levels to the levels in *Sox2eGFP* cells were determined. Statistical analysis was performed using the paired t-test, except for flow cytometry results (Wilcoxon signed-rank test) and smFISH results (Mann-Whitney U test).

Sox2otb/c RNA levels were rather negatively correlated with *Sox2* RNA levels (Fig. 1f) but the heterogeneic nature of the cultures thwarts to directly link *Sox2otb/c* levels to *Sox2* levels. To investigate whether *Sox2* levels are indeed negatively correlated with *Sox2otb/c* levels we measured the levels of *Sox2* and *Sox2otb/c* in *Sox2* heterozygous and homozygous ESC lines, in multiple monoclonal ESC-derived, radial glia-like neural stem (NS) cell lines generated from wild type mouse ESCs, and in neurosphere cultures of primary NPCs from the lateral ventricle of the adult mouse brain. NS cells express two to three-fold less *Sox2* RNA[44,45] (Fig. 1g) but contain higher levels of *Sox2otb/c* RNA in comparison with mouse ESCs. Primary NPCs contain higher *Sox2otb/c* RNA levels, whereas *Sox2* levels are further reduced (Fig. 1g). In contrast to previous studies on *Sox2ot* expression in immortalized transformed cells[32–34], we observed a negative correlation between *Sox2otb/c* and *Sox2* RNA levels (Spearman r = −0,7857, P-value = 0.048)(Supplementary Fig. S2h).

Transcriptional activity of Sox2ot alters Sox2 RNA levels in cis

Next we wondered whether the negative correlation between *Sox2ot* and *Sox2* is caused by a direct mechanism. Long non-coding RNAs are known to regulate neighboring genes in a variety of ways by either a *cis* (only the allele from which the lncRNA is transcribed is affected) or *trans* (the effect is independent of the allele from which the lncRNA is transcribed) mechanism. However, knocking out all three *Sox2ot* genes (*Sox2otb*, *Sox2otc*, and the 545 kb upstream of *Sox2* located *Sox2dot* (Supplementary Fig. S1b)) simultaneously is extremely difficult. Moreover, such a strategy would likely perturb ordinary locus regulation as

removal of critical *Sox2ot* promoter sequences may delete important regulatory sequences that are key for correct expression of neighboring genes. To circumvent these pitfalls, we opted to enhance the transcriptional activity of *Sox2otb* in *Sox2* expressing cells that normally contain very low levels of *Sox2ot*. We introduced the human ubiquitin C (*UbiC*) promoter directly upstream of *Sox2otb* exon 1 by homologous recombination in mouse *Sox2eGFP* ESCs (Fig. 2a,c), which have one copy of *Sox2* replaced by eGFP[22].

Three clones contained an insertion of the UbiC promoter into the *eGFP* allele (*UbiCeGFP*) and two into the *Sox2* allele (*UbiCSox2*) (Fig. 2d, e). *Sox2otb* was highly transcribed in all targeted ESCs, albeit, levels were lower when the UbiC promoter was inserted into the *Sox2* allele, hinting towards the existence of an allele-specific modulatory mechanism (Fig. 2f). If the negatively correlated expression of *Sox2* and *Sox2ot* is an immediate consequence of *Sox2ot* expression, an effect on *Sox2* as well as eGFP (trans regulation) or, on either *Sox2* or eGFP (cis regulation) should be evident in the targeted cells. Indeed, *Sox2ot* transcription resulted in a 20–30% reduction in *Sox2* or eGFP levels (Fig. 2g, h). However, reduced expression was solely observed for the gene (*Sox2* or eGFP) that was located on the targeted allele. These data demonstrate that *Sox2ot* transcription regulates *Sox2* transcription in cis. Although reductions were relatively moderate, a compensatory mechanism was activated in the ESCs that have decreased *Sox2* levels as illustrated by enhanced eGFP levels. This is reminiscent of the results in hybrid ESCs, in which allele-specific reduction of *Sox2* by deletion of the ESC prevalent transcriptional enhancer led to upregulation of *Sox2* from the unmodified allele[20].

To determine whether the *Sox2* downregulation is specific for the whole population or whether only a proportion of the population contributed to the lower *Sox2* levels we quantified *Sox2* RNA at the single cell level by smFISH. smFISH allows us to count the expression of individual RNA molecules in individual cells, which reveals expression heterogeneity within the population. We measured *Sox2* levels in 700 cells of each ESC line (Fig. 2i, and Supplementary Fig. S3b). Only when *Sox2ot* was expressed from the *Sox2* allele we observed a ~20% reduction in the means (77 versus 96 (*Sox2eGFP*) or 98 (*UbiCeGFP*) transcripts). Moreover, the distribution of *Sox2* gene expression in *UbiCSox2* cells differed from *UbiCeGFP* and the parental *Sox2eGFP* cells (Mann-Whitney U;

FDR = 3.19e-10 and FDR = 1.11e-10, respectively), whereas the distributions in *UbiCeGFP* and *Sox2eGFP* cells were comparable. This analysis confirmed that *Sox2* RNA levels are decreased when *Sox2ot* is transcribed from the same allele and showed that this effect is likely not restricted to a subpopulation of cells (Fig. 2i).

Mouse ESCs overexpressing Sox2ot are very similar to wild type ESCs

Next we investigated the effect of *Sox2ot* overexpression on the maintenance and differentiation of mouse ESCs. On basis of morphology we could not identify phenotypic differences between the parental *Sox2eGFP* ESCs and the *Sox2ot* overexpressing ESCs (Fig. 3a). The absence of a maintenance phenotype was underscored by the analysis of the expression of platelet endothelial cell activation marker CD31 (PECAM) and stage-specific embryonic antigen (SSEA1), which discriminates naïve and primed pluripotent cell states[44,45]. *Sox2eGFP* and *Sox2ot* overexpressing lines showed a similar and homogeneous CD31 expression profile, whereas SSEA1 was more heterogeneously expressed which is a normal feature of ESCs (Fig. 3b). Also the expression of other pluripotency genes like Nanog and Oct4 was not altered (Supplementary Fig. S4a, S4b). In addition, prolonged passaging at a constant splitting ratio did not reveal gross differences in the expansion rate between *Sox2eGFP* and *Sox2otb* overexpressing ESCs (data not shown). Possibly this is due to adaptation of the *UbiCSox2* ESCs to lower levels of *Sox2* RNA by acquiring more normal SOX2 protein levels (Supplementary Fig. S4c). Since *Sox2otb* is induced during the differentiation of ESCs into neuroectoderm we also investigated the effect of *Sox2otb* overexpression on neuroectodermal differentiation. Using EB-based differentiation protocols we could not detect quantitative or temporal differences in the generation of either NSCs or more mature Tubb3 positive cells (Fig. 3c,d). In addition, the differentiation into mesendoderm as determined by Brachyury expression is largely unaltered (Supplementary Fig. S4d). Taken together these results indicate that enhanced *Sox2ot* levels do not majorly alter the phenotype of ESCs and do not exert gross effects on the EB-based differentiation of mouse ESCs.

Sox2otb/c is enriched in the nucleus but not associated to chromatin

Many lncRNAs that regulate transcription are enriched in the nucleus. We therefore investigated the cellular localization of *Sox2ot*. As our *Sox2ot* exonic smFISH probe set was not specific enough, we analyzed the cellular localization of

Sox2ot RNA by cell fractionation and qRT-PCR. *Sox2ot* RNA was 4 times more enriched in the nucleus than *Sox2* RNA but 6 times less than Neat1, a lncRNA that is highly abundant in the nucleus[46] (Fig. 4a).

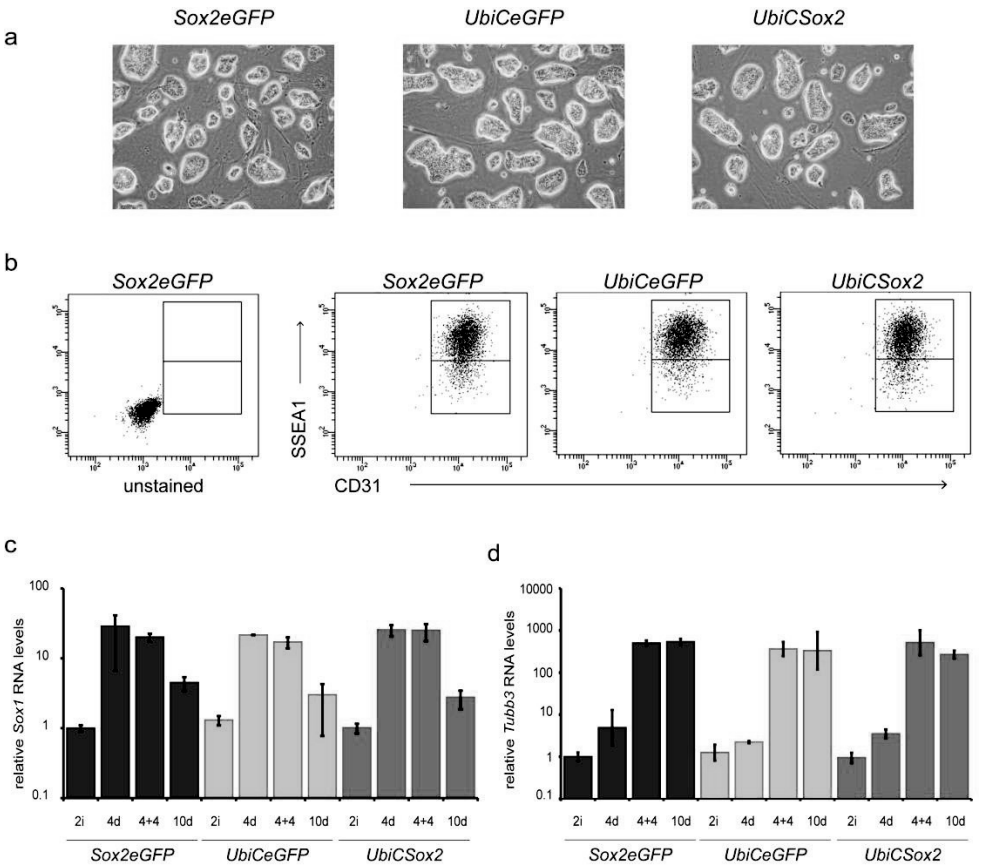

Figure 3. ESCs overexpressing endogenous *Sox2otb* are similar to *Sox2eGFP* ESCs. (a) Phase contrast pictures of *Sox2eGFP*, *UbiCeGFP*, and *UbiCSox2* cells cultured in 2i medium (100x magnification). (b) SSEA1 and CD31 expression in *Sox2eGFP*, *UbiCeGFP*, and *UbiCSox2* cells as measured by flow cytometry. (c and d) RNA levels of *Sox1* (c) and Tubb3 (d) during EB-mediated neural differentiation of *Sox2eGFP*, *UbiCeGFP*, and *UbiCSox2* cells as measured by qRT-PCR. RNA levels were normalized against β-actin. RNA levels relative to the levels in *Sox2eGFP* cells are depicted on a 10 log scale. The results of one representative experiment (out of three independent experiments) using (sub)clones of *Sox2eGFP* (n = 2), *UbiCeGFP* (n = 3), and *UbiCSox2* (n = 2) is depicted as mean +/− SD.

Next we examined whether *Sox2ot* is associated to the chromatin fraction. LncRNAs that function through a *trans*-acting mechanism are often found

enriched in the chromatin fraction, like Neat1 [46]. In support of the observed *in cis* effect of *Sox2otb/c* we predominantly found *Sox2ot* RNA in the soluble nuclear fraction (Fig. 4b).

H3K4 methylation is unaltered in Sox2otb overexpressing mouse ESCs

The allele-specific regulation of *Sox2* prompted us to investigate the nature of this regulation. A large group of cis-acting lncRNA transcripts represses genes by recruiting chromatin-modifying proteins that install a repressive histone mark such as H3K27me3 or H3K9me3, or by controlling H3K4 methylation[47]. To gain evidence for the existence of a *Sox2ot* dependent chromatin-modifying mechanism we compared H3K4me1, H3K4me2, H3K4me3, H3K9me3, and H3K27me3 chromatin marks in the region between the first exon of *Sox2otb* and the last exon of *Sox2otb/c* in cells expressing *Sox2* and *Sox2ot* at different ratios, i.e. ESCs and ESC-derived NPCs, using publicly available H3 methylation chromatin immunoprecipitation-sequencing (ChIP-seq) data sets (Fig. 4c, and Supplementary Fig. S5a and d). The only histone methylation profiles that are strongly altered between ESCs and ESC-derived NPCs are confined to a conserved region in the proximal enhancer/promoter region of *Sox2* ~4 kb downstream of the first exon of *Sox2otc* (Supplementary Fig. S5a,b, and c). In this region H3K27me3 and H3K4me3 were high in ESCs indicating a bivalent chromatin signature, which is linked to key developmental genes[48,49]. The bivalent histone status is lost in this region in ESC-derived NPCs. We wondered whether overexpression of *Sox2otb* would change the ESC chromatin into a more NPC-like chromatin regarding H3K4me3. We performed H3K4me3 ChIP assays for this region but did not observe differences in H3K4me3 between the cell lines (Fig. 4d). Although we did not rule out the involvement of other epigenomic changes, we decided to investigate other candidate regulatory mechanisms.

Sox2otb transcription impairs the formation of the chromatin promoter-enhancer loop driving expression of Sox2

Development and homeostasis require coordinate regulation of neighboring genes through enhancers and locus control regions[50]. Chromatin looping enables transcription activation by juxtaposing locus control regions (LCRs), distal regulatory elements and promoter elements, and thus, function by bringing transcription factors, coactivators, and RNA polymerase II together. In ESCs multiple chromatin loops exist in the *Sox2* locus[51]. The most prevalent chromatin

interaction is formed by the *Sox2* regulatory region 1 (SRR1) upstream of *Sox2* and a 13 kb super enhancer termed *Sox2* control region (SCR) located ~100 kb downstream of *Sox2* (Fig. 4e)[20,52]. Deletion of this super enhancer decreases *Sox2* levels in mouse ESCs 6 to 9 fold[20,52]. Thus, if a decrease in *Sox2* levels were the consequence of *Sox2otb* mediated transcriptional interference the SRR1-SCR interaction would likely be diminished.

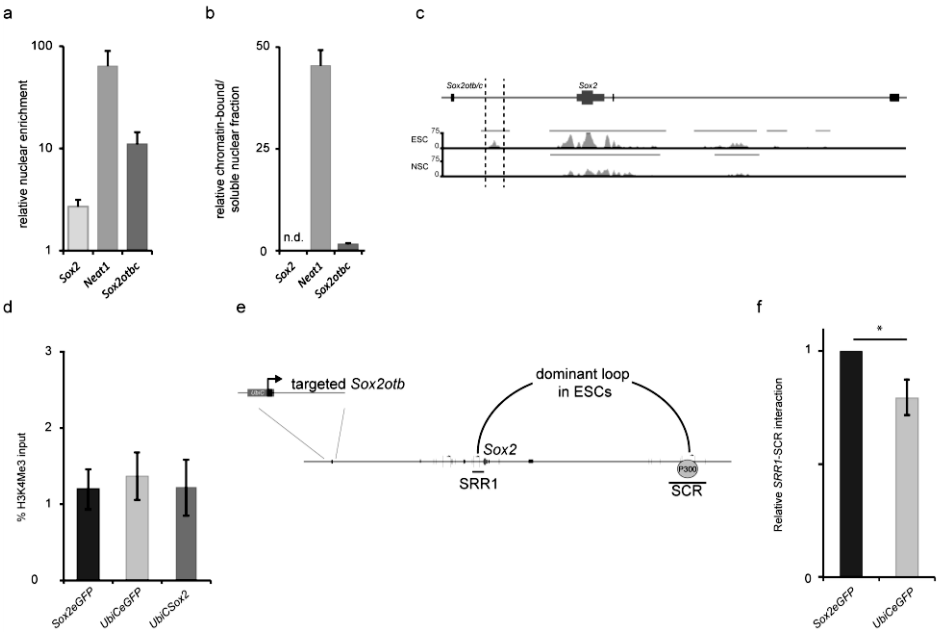

Figure 4. *Sox2* locus-specific H3K4 trimethylation and chromatin interactions in ESCs overexpressing *Sox2otb*. (a) Analysis of *Sox2ot* RNA localization in ESCs. *Sox2ot* is enriched in the nucleus when compared to β-Actin as determined by qRT-PCR after subcellular fractionation. The ratio (+SD) of nuclear/total RNA (200ng input) relative to that of β-actin is depicted on a 10 log scale. Neat1 is a lncRNA that is enriched in the nucleus, and which is predominantly associated to chromatin[46]. (b) Analysis of the nuclear localization of *Sox2ot* in ESCs by qRT-PCR. The depicted ratio of chromatin bound RNA (+SD) is relative to that of β-actin. (c) Genome browser view of H3K4me3 density signals in the regulatory *Sox2* region of ESCs and ESC-derived NPCs[67]. For quantification of the difference see Supplementary Fig. 5Sa, b, and c. (d) H3K4me3 ChIP results for the region depicted between vertical lines in (c). Depicted H3K4me3 levels are relative to H3K4me3 levels of the housekeeping gene Myl6. (e) Schematic drawing of the dominant chromatin loop in ESCs formed by interaction of the *Sox2* proximal region (*Sox2* regulatory region 1) (SRR1) with a P300 bound super enhancer (SCR) located ~110 kb downstream of *Sox2*. HindIII fragments and primers used are shown. (f) 3C chromatin conformation capture of the SRR1-SCR interaction depicted in (e). Values are relative to interactions of the *Sox2* intergenic region upstream of *Sox2otc*. Values are represented as mean +/− SD from three independent experiments (n = 10). *Paired t-test P value = 0.02.

Through chromosome conformation capture (3C) we analyzed whether the SRR1-SCR chromatin interaction was altered in *Sox2otb* overexpressing (*UbiCeGFP*) ESCs, which did not show altered *Sox2* levels, compared to parental *Sox2eGFP* ESCs. We indeed observed a lower frequency of SRR1-SCR interactions in *Sox2otb* overexpressing cells versus *Sox2eGFP* cells (Fig. 4f). In summary, transcriptional activity of *Sox2otb* negatively correlates with *Sox2* levels, and in addition, enhanced *Sox2otb* transcription correlates with reduced chromatin interactions between the upstream regulatory sequence of *Sox2* and the super enhancer of *Sox2* in mouse ESCs.

Discussion

Through transcription analysis in combination with genetic modification of the endogenous *Sox2otb* locus we identified that transcriptional activity of *Sox2otb* represses *Sox2* expression in mouse ESCs. In contrast to our findings, previous studies in human cancer as well as cancer cell lines have demonstrated a positive correlation between *Sox2ot* and *Sox2* in certain but not all cell types investigated[32–34]. A quantitative and qualitative comparison of the published expression data is rather difficult due to the genomic positions of the primers used as the applied primer pairs recognize either a variety of *Sox2ot* splice variants or amplify only *Sox2ot* sequences downstream of *Sox2*. Nevertheless, the positive co-regulation of *Sox2* by *Sox2ot* has been strongly supported by ectopic overexpression or knockdown of *Sox2ot* pointing to a trans effect[32–34]. One may argue that transcription regulatory mechanisms of certain genes in human cells are different from those in murine cells, however, the strong conservation of the whole *Sox2ot* genomic region rather suggests a highly similar mode of operation. We believe that the disparities with the results obtained in this study are more likely caused by the differences in the cells analyzed, as gene regulation is very much cell type specific. In addition, cancer cells have undergone many epigenetic and genetic changes that interfere with the specificity and integrity of regular gene transcription programs[53]. Since we investigated early neural development using non-transformed mouse cells our data indicate that *Sox2* regulation during stem cell maintenance and differentiation is completely different from *Sox2* regulation in cancer cells.

Cis regulation of neighboring genes has been proposed to be an important function of many lncRNA genes, but up to now this has only been proven for a

very small subset of lncRNAs predominantly involved in imprinting and X inactivation because of the more easily detectable allele-specific modifications[47]. In general, a major hurdle has been selection of an allele-specific genomic modification strategy to identify allele-specific differences that represent a bona fide phenotype. In addition, modification of lncRNA genes to study cis-acting mechanisms is rather challenging. Introducing single or small mutations that alter the function or expression of uncharacterized lncRNAs is very complicated due to the non-coding nature of these genes. Nevertheless, insertion of a strong polyadenylation signal that prematurely truncates the lncRNA transcript has been successfully exploited to gain insight into the requirement of the full-length lncRNA[27]. However, premature polyadenylation strategies do not allow analysis of the role of lncRNA transcription initiation or that of promoter/enhancer sequences. Instead deletion of presumed important regulatory regions may be considered to address their role. Recent genome editing advances using CRISPR/Cas9 have facilitated the deletion of genomic sequences[54] but deletion of important promoter or exon sequences imposes the risk of removing important transcriptional regulatory regions of the neighboring genes, in particular, because lncRNAs are often transcribed from enhancer and promoter proximal sequences of adjacent genes. This may result in false attribution of the role of the modified lncRNA. As to *Sox2ot* the existence of at least three independent transcriptional initiation sites of *Sox2ot*, and possibly more as indicated by human CAGE datasets, would make the generation of a full knockout rather unrealistic. Moreover, one of the *Sox2ot* transcription initiation sites (that of *Sox2otc*) is located in regulatory sequences proximal of *Sox2*. Deletion of this genomic sequence may directly alter *Sox2* transcription independent of *Sox2otc*. As feasible alternative we created a promoter insertion that drives transcription of only one of the *Sox2ot* genes to study the role of *Sox2otb* overexpression in development and the regulation of *Sox2*. Using this overexpression system we demonstrate that the reduction in *Sox2* RNA levels is caused by allele-specific transcriptional activity of *Sox2otb*. The reduced levels of *Sox2* did not exert a loss of pluripotent stem cell self-renewal phenotype, as may have been expected, likely due to adaptation of the ESCs to decreased *Sox2* RNA levels. It is known that a decrease in *Sox2* levels in ESCs activates a feedback mechanism enhancing expression of *Sox2* [20]. Also in the *UbiCSox2* cells we observed upregulation of the other *Sox2* allele (here *eGFP* allele) indicating the activation of such feedback loop and the importance of having higher levels of *Sox2*. However, since the

other allele is non-functional, enhanced expression of the other allele was ineffective. Instead the *UbiCSox2* cells adapted to lower *Sox2* levels by regaining SOX2 to a level similar to that of the parental *Sox2eGFP* cells.

Sox2 is also crucial for neuroectodermal differentiation of ESCs, and lower *Sox2* levels favor mesendoderm commitment[9]. If the SOX2 protein levels would not have been enhanced upon adaptation a differentiation phenotype would have been expected in the cells that overexpress *Sox2otb* from the *Sox2* allele. Although *Sox2* adaptation may have obscured an early neuroectodermal, Sox2-dependent differentiation defect, a Sox2-independent trans effect was not observed. Thus our results indicate that the main function of *Sox2otb* is cis regulation of *Sox2* rather than affecting cell physiology in trans via other routes. The importance of *Sox2ot* transcriptional activity is underscored by the genomic conservation of *Sox2ot* between mammals and *fugu*, which is much higher in *Sox2ot* promoter (proximal) sequences than exon sequences (Supplementary Fig. S1c).

The introduced *Sox2otb* transcriptional activity led to decreased *Sox2* transcription and reduced interaction of the *Sox2* proximal promoter region (SRR1) with the ESC-specific enhancer in this genomic region. However, we cannot rule out that other chromatin interactions are affected as well. In the presented heterozygous ESC model maximally 50% of a specific chromatin loop can be altered when considering an in cis effect. Therefore, only differences in very dominant chromatin loops, either the ones that are newly formed or the regular ones, are detectable. A hypothetical mechanism that would fit our observations is transcriptional repression by virtue of blocking recruitment of RNA polymerase II to the SSRI region (Fig. 5a). A very similar mechanism is exploited by Airn, which repress Igf2r by preventing RNA polymerase II recruitment to the Igf2r promoter[27]. As well-balanced *Sox2* protein levels are crucial for correct development of the distinct subsets of neurons in the neural tube[10,11], it is tempting to speculate that during development the main function of *Sox2ot* is controlling *Sox2* levels. In this respect the 20–30% reduction in expression of *Sox2* RNA that we have observed may seem irrelevant. However, recently it became clear from single cell RNA sequence analysis in primary mouse cortical NSCs/NPCs that *Sox2* dosage regulates their division rate and controls their ability to maintain an undifferentiated state[55]. This study demonstrated that

very modest decreases in *Sox2* levels in NSCs/NPCs are accompanied by rapid increments of the neuronal specification factor Neurogenin2 (Fig. 5b). Moreover, an approximately 20% reduced expression of *Sox2* (the same reduction as we observed in *Sox2ot* overexpressing ESCs) appeared to be a threshold for expression of the neural differentiation markers Eomes and Tbr1 (Fig. 5b). These data indicate that a subtle decrease of *Sox2* may have a profound impact on the status of NSC/NPCs regarding their differentiation potential, and that *Sox2ot* transcription through the *Sox2* gene may render NSCs/NPCs more susceptible to neural differentiation.

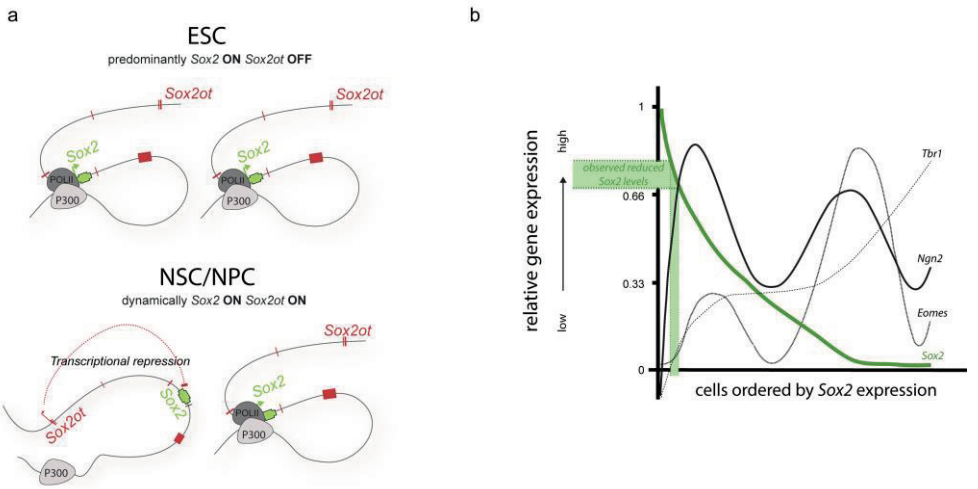

Figure 5. Proposed model of transcriptional interference to modulate *Sox2* levels during neural development. (a) Hypothetical model illustrating *Sox2* and *Sox2ot* transcription in ESCs and NSCs/NPCs. In ESCs *Sox2* (green) is predominantly transcribed, whereas *Sox2ot* (red) transcription is predominantly off. In NSCs *Sox2* and *Sox2ot* are transcribed in the same cell. On basis of our data we propose a dynamic on/off situation. If *Sox2ot* is transcribed *Sox2* transcription is repressed, and vice versa. (b) Adapted graph from a single cell RNA profiling study by Hagey and Muhr[55] showing the influence of subtle reductions in *Sox2* on the expression of neuronal genes. The observed 20–30% reduction in *Sox2* transcription by *Sox2ot* transcription (indicated by green bars) lies at the threshold of the expression of neural genes Tbr1 and Eomes and corresponds to a steep rise in the expression of the proneural gene Ngn2 in cortical NSCs/NPCs.

We believe that the here proposed role of *Sox2ot* is likely conserved in numerous loci containing key differentiation genes. Transcriptome data have revealed that analogous overlapping transcripts are present in the *Sox1* and *Sox4* loci. It will be

interesting to learn the underlying nature of these Sox regulatory mechanisms, to what extent this regulation exists in the mammalian genome, and how disruptions disturb development.

Methods

Cell culture

Mouse ES cell lines (E14-cl22[44], E14 subclone IB10, R1, CCE, and *Sox2eGFP*[22] (parental mouse ESCs as well as the targeted clones) were cultured feeder-free or on irradiated mouse embryonic fibroblasts (MEFs) on 0.1% gelatinized tissue culture surface in DMEM containing 1 mM L-glutamine, 1x non-essential amino acids (NEAA), PenStrep 1%, 1000 U/ml human LIF (Peprotech), 50 µM β-mercaptoethanol and 15% mouse ESC tested fetal bovine serum (FBS) (Life Technologies). *Sox2eGFP* ESCs were kindly provided by the late Dr. L. Pevny, University of North Carolina, Chapel Hill. For ChIP, 3C, differentiation and qRT-PCR cells were first cultured for 4 passages in 2i medium[56] (DMEM/F12/NeuralBasal, Glutamax, PenStrep 1%, human LIF 1000 U/ml (Peprotech), 50 µM β-mercaptoethanol, 0.5x B27 plus vitamin A (ThermoFisher), 0.5x N2 (ThermoFischer), 1 µM PD0325901 (Axon Medchem) and 3 µM CHIR99021 (Axon Medchem) and a FBS percentage that was gradually decreased from 15% to 1%. Cells were passaged using Trypsin/EDTA (0.05%/0.02%). Cells were maintained at 37 °C and 5% CO2. Prior to the experiments the quality of the cells was analyzed by flow cytometric analysis using anti-mouse SSEA1-BV421 (BD) and anti-mouse CD31-PerCPefluor710 (eBioscience) antibodies. SOX2 was measured by flow cytometry using a goat anti-Sox2 polyclonal antibody (SantaCruz, Biotechnology, sc-17319), in combination with an anti-goat-Alexa568 secondary antibody (Thermo Fisher Scientific). Staining was performed usingthe fix & perm kit (Thermo Fisher Scientific) according the manufacturer's instructions.

Targeting Sox2eGFP mouse ES cells

Two independent homologous recombination experiments were performed using *Sox2ot*b targeting vectors containing UbiCloxPHyTKpAloxP or UbiClox2272PurDTKpAlox2272 selection cassettes. The selection modules were inserted 9 nucleotides upstream of the identified *Sox2otb* transcription start site (chr 3: 34,459,297 NCBI37/mm9) into the genomic sequence (chr 3: 34,453,460–

34,463,055 NCBI37/mm9) that was amplified from 129Ola genomic DNA using Phusion polymerase (NEB). The knockin constructs were introduced into *Sox2eGFP* ESCs by electroporation, and drug resistant clones were selected using hygromycin (110 µg/ml) or puromycin (1.5 µg/ml). Homologous recombinants were identified by Southern blot analysis of EcoRV restricted genomic DNA using ^{32}P labelled *Sox2otb* flanking probes. In total, 465 colonies were screened for correct homologous recombination. Five correctly recombined clones were further investigated to identify whether the *Sox2* or *eGFP* allele was targeted. To this end SbfI restricted genomic DNA was separated by pulse field gel electroforesis (PFGE) and analyzed by Southern blotting using ^{32}P labelled eGFP and *Sox2* probes. Three clones contained an insertion of the UbiC promoter into the *eGFP* allele (*UbiCeGFP*) and two into the *Sox2* allele (*UbiCSox2*).

RNA in situ hybridization

Whole mount in situ hybridization was performed according standard protocols. In short, dissected E9.25 embryos (C57Bl/6) were fixed in 4% PFA O/N. Fixed embryos were twice washed in PBS 0.1% Tween-20 (Sigma) (PBST), and dehydrated by subsequent methanol washing steps (25-50-75 and 100% methanol). Dehydrated embryos were slowly rehydrated (10' per step) at RT while rotating. After rehydration the brain vesicle was punctured and the surround membrane ruptured to prevent trapping of the riboprobes. Embryos were treated with proteinase K (10 mg/ml) for 10', and gently rinsed in PBST. Next embryos were again fixed in 4% PFA and 0.2% glutaraldehyde for 20' while rotating, washed in PBST, and incubated in 50% PBT/50% hybridization solution (HS) (HS: 50% formamide (Sigma), 1.3x SSC, pH 5.0 (Ambion), 5 mM EDTA, pH 8.0 (Ambion), 50 mg/ml yeast tRNA (Sigma), 0.2% Tween-20 (Sigma), 0.5% CHAPS (Thermo Fisher Scientific), and 100 mg/mL Heparin (Sigma)), and subsequently 100% HS. Riboprobes, generated by T7 polymerase in vitro transcription (antisense and sense *Sox2*, *Sox2otb* and *Sox2OTb/c* digoxigenin labeled RNA probes (sequences in Table S1)), were added to HS and incubated for 20 hours at 70 °C. Embryos were washed 3 times with 2x SSC, 0.1% CHAPS, three times with 0.2x SSC, 0.1% CHAPS, and twice with 1x KTBT (50 mM TrisHCl, pH 7.5, 150 mM NaCl, 10 mM KCl and 1% Triton X-100). Embryos were incubated with 10 ug/ml of RNase A in KTBT for 30 min. at 37 °C, blocked with 2% blocking solution (Roche), and 20% heat inactivated sheep serum, and subsequently O/N incubated with AP conjugated a-DIG, Fab fragment (Sigma) in the same blocking buffer at 4 °C.

Embryos were 5 times washed in 0.1% Tween-20 and 1 mM levisamole (Roche) in ddH2O, and subsequently stained in 1x BM purple (Roche) plus 0.1% Tween, 1 mM levamisole. Reaction was stopped by washing in ddH2O. Whole mount stained embryos, were embedded in 2% agarose and cross-sectioned on a vibratome (Leica). Mice were maintained under specific-pathogen-free conditions. All animal experiments were approved by the Animal Experiments Committee of the LUMC performed to the recommendations and guidelines set by the LUMC and by the Dutch Experiments on Animals Act that serves the implementation of guidelines on the protection of experimental animals by the Council of Europe.

RNA-linker mediated (RLM)-RACE and in vitro transcription translation

The used RLM-RACE procedure has been extensively described elsewhere[57]. *Sox2ot* reverse primers were located in exon 1 of *Sox2otc*. In vitro transcription/translation of human TP53 and the full-length *Sox2otb* and *Sox2otc* cDNA sequences was performed using TNT® Quick Coupled Transcription/Translation System (Promega) according the manufacturer's protocol. ^{35}M labeled proteins were separated on 5–15% and 20% polyacrylamide gels.

ESC differentiation

For embryoid body (EB) differentiation the original protocol was slightly adapted[58]. For neural differentiation: ESCs were seeded as a single cell suspension at a concentration of 100,000-200,000 cells/ml in ESC media containing FBS (as in the original protocol) or knockout serum replacement (KSR)[59] lacking hLIF and 2i on ultra-low attachment plates (Corning). After 4 days of culture all trans retinoic acid (ATRA) (Sigma) or the synthetic substitute EC23 (Abcam) was added to the media at a concentration of 0.5 µM. Media was changed once every two days. For mesendodermal differentiation, aggregated ESCs were cultured in 2i media containing 3 µM CHIR99021 but without PD0325901, hLIF, and FBS as has been described for monolayer differentiation[9]. 3 ½ days after addition of CHIR99021 EBs were manually dissociated using the embryoid body dissociation kit (Miltenyi Biotech) according the manufacturer's instructions. Cells were stained for Oct4 and Brachyury using mouse anti-Oct4-BV421 (BD) and goat anti-Brachyury (SC-17745, SantaCruz) and a secondary donkey anti goat Alexa568 antibody (Thermo Fisher Scientific) using the fix &

perm kit (Thermo Fisher Scientific) according the manufacturer's protocol. Oct4 and Brachyury expression was measured on a LSRII flow cytometer (BD). For monolayer differentiation we adapted the protocol used by Engberg *et al*.[60]. In brief, mouse ESCs were seeded at a density of 15,000 cells/cm^2 onto 0.1% gelatin (Sigma) coated dishes in 2i media, lacking hLIF and 2i, but containing 1% FBS. Media was replaced with DMEM/F12/Neuralbasal containing L-glutamine (ThermoFisher), PenStrep 1%, 1x N2 (ThermoFisher), and 1x B27 without vitamin A (ThermoFisher), and ATRA (Sigma) or EC23 (Abcam), or hBMP4 (Peprotech) at the concentrations indicated, 12 hours after seeding the cells. Cells were cultured for the indicated periods and media was replaced every two days. NS cell lines were generated from different ESC lines using N2B27 media as described elsewhere[61]. One of the clones has been extensively characterized[44].

RNA isolation and quantitative PCR analysis

Total RNA was isolated directly from the cells using Trizol (Life technologies) or NucleoSpin® columns (Macherey-Nagel). Following DNaseI treatment (Roche), cDNA was generated from 100–500 ng RNA using Transcriptor reverse transcriptase (Roche) and random hexamers or an oligod(T) primer according the manufacturer's protocol. After the samples had been checked for genomic DNA contaminations, cDNA was measured quantitatively on a Bio-Rad CFX96 using SensiFASTTM Sybr green PCR mix (Bioline) and the primers listed in Supplementary Table S1. All primers were tested for a comparable and linear amplification efficiency using a dilution series of cDNA or gDNA. RNA levels were normalized against β-actin and 18 S, which yielded similar outcomes. For direct quantitative comparison of expression levels between ESCs and NS cells levels were normalized against housekeeping gene Myl6 because Myl6 expression is unaltered between ESCs and NSCs[44]. All measurements were performed in triplicate. Relative expression was calculated using the comparative Ct method, known as the 2–[delta][delta]Ct method, where [delta][delta]Ct = [delta]Ct(sample) - [delta]Ct. Dependent on the experiment, the reference samples were the 2i samples (also described as day 0 of differentiation), or the parental ESC line *Sox2eGFP*.

Single molecule fluorescence in situ hybridization (smFISH)

Mouse ESCs were cultured in 2i medium or differentiated in N2B27 media without additives for 4 days as described above. Cells were detached with

Accutase (Gibco), resuspended in serum containing medium, and fixed by adding paraformaldehyde to an end-concentration of 4% and subsequent incubation for 12 minutes at RT. Fixed cells were pelleted by a 3' centrifugation and subsequently resuspended in 70% ethanol. Samples were stored at 4 °C until use. smFISH of *Sox2* (Stellaris VSMF-3075-5-BS probe set) was performed exactly as before[62] and signals were quantified using custom MATLAB scripts. *Sox2ot* transcription was determined using a custom probe set covering *Sox2otb* intron 2, which was designed by homemade MATLAB scripts.

Chromatin immunoprecipitation (ChIP) and 3C conformation capture

The chromatin of a single cell suspension of mouse ESCs was crosslinked in ESC medium containing 1% formaldehyde. Protocols used were previously described by Lee *et al*.[63] (ChIP) and Stadhouders *et al*.[64] (3C). For ChIP: the nuclear fraction was sonicated for 9 minutes (30" on, 30" off) using a Biorupter UCD-200 (Diagnode). After sonication, H3K4me3 chromatin was precipitated overnight at 4 °C in 0.1% fraction V BSA, protease inhibitors (Roche), 16.7 mM trisHCl, 167 mM NaCl, 1.25 mM EDTA, 0.01% SDS, 1% Triton X-100, Dynabeads Protein G (ThermoFisher) and 1 ug H3K4me3 rabbit polyclonal antibody (Diagenode). Chromatin was eluted in 1% SDS, and 0.1 M NaHCO3, de-crosslinked at 65 °C for 8–12 hours, treated with RNAse A and ProtK, and purified using phenol/chloroform extraction. Mouse insulin promoter primers and Myl6 primers were used as negative control and positive/normalization control, respectively. For 3C: chromatin was restricted with HindIII (Fermentas) for 24 hours and O/N ligated at 16 °C. Chromatin was de-crosslinked at 65 °C for 8–12 hours, treated with RNAse A and ProtK, and purified using phenol/chloroform extraction. Quality and quantity of DNA was checked by a linear amplification of Sox2UTR genomic sequences. Ligation efficiencies were checked through amplification of ESC-specific Dppa2 chromatin loop[65].

Subcellular fractionation

Cell fractionation: mouse ESCs were divided into two fractions and used for either total RNA isolation or nuclear RNA isolation. Nuclear RNA was isolated as previously described[57]. In brief, cells were lysed and nuclei were pelleted after centrifugation (1350 *g* at 4 °C for 5 min). Cells (total RNA) or nuclei (nuclear RNA) were lysed using RA1 RNA lysis buffer (Macherey-Nagel) and RNA was isolated on NucleoSpin® columns (Macherey-Nagel) according to the manufacturer's

instructions. 200 ng of RNA was used in the reverse transcription reaction that was performed as described above.

Nuclear fractionation: Fractionation of the nucleus was performed as described by Werner *et al*.[66] In brief, crude nuclei were resuspended in 250 µl NRB (20 mM HEPES pH 7.5, 50% Glycerol, 75 mM NaCl, 1 mM DTT, 1x protease inhibitor cocktail) and centrifuged for 5′ at 500 *g* at 4 °C. The pellet was again resuspended in 250 µl NRB and 1 volume of NUN buffer ((20 mM HEPES, 300 mM NaCl, 1 M Urea, 1% NP-40 Substitute, 10 mM MgCl2, 1 mM DTT) was added, followed by a 5′incubation on ice after which the suspension was centrifuged (1,200 g, 5 min, 4 °C). The soluble fraction supernatant was transferred to a tube and the pellet was resuspended in 1 ml buffer A (20 mM HEPES pH 7.5, 10 mM KCl, 10% glycerol, 340 mM sucrose, 4 mM MgCl2, 1 mM DTT, and 1x Protease Inhibitor Cocktail (Roche) and centrifuged (1,200 *g*, 5 min, 4 °C). The chromatin pellet was resuspended in 50 µl buffer A, and 500 µl Trizol (Life technologies) was added. Trizol was added as well to the soluble nuclear fraction. Subsequently, RNA was extracted following the manufacturer's guidelines. cDNA was generated as described above.

Supplementary information

Supplementary information is available online on the website of Scientific reports (dx.doi.org/10.1038/s41598-017-18649-4).

Acknowledgements

This study was supported by the Dutch Brain foundation and ZonMW. We thank the late L. Pevny (University of North Carolina, Chapel Hill) for the *Sox2eGFP*cells, the LUMC sequence analysis support core for ChiPseq data alignments, and S. Chuva de Sousa Lopes (LUMC) for assessing the ISHs. The authors declare to have no competing financial interests.

References

1. Kamachi Y, Kondoh H. Sox proteins: regulators of cell fate specification and differentiation. Development. 2013;140:4129–4144. doi: 10.1242/dev.091793.
2. Nichols J, Smith A. Pluripotency in the embryo and in culture. Cold Spring Harb Perspect Biol. 2012;4:a008128. doi: 10.1101/cshperspect.a008128.
3. Ivanova N, *et al*. Dissecting self-renewal in stem cells with RNA interference. Nature. 2006;442:533–538. doi: 10.1038/nature04915.

4. Kopp JL, Ormsbee BD, Desler M, Rizzino A. Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. Stem Cells. 2008;26:903–911. doi: 10.1634/stemcells.2007-0951.

5 Masui S, *et al*. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol. 2007;9:625–635. doi: 10.1038/ncb1589.

6 Li J, *et al*. A dominant-negative form of mouse SOX2 induces trophectoderm differentiation and progressive polyploidy in mouse embryonic stem cells. J Biol Chem. 2007;282:19481–19492. doi: 10.1074/jbc.M702056200.

7 Fong H, Hohenstein KA, Donovan PJ. Regulation of self-renewal and pluripotency by Sox2 in human embryonic stem cells. Stem Cells. 2008;26:1931–1938. doi: 10.1634/stemcells.2007-1002.

8 Adachi K, Suemori H, Yasuda SY, Nakatsuji N, Kawase E. Role of SOX2 in maintaining pluripotency of human embryonic stem cells. Genes Cells. 2010;15:455–470.

9 Thomson M, *et al*. Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell. 2011;145:875–889. doi: 10.1016/j.cell.2011.05.017.

10 Bylund M, Andersson E, Novitch BG, Muhr J. Vertebrate neurogenesis is counteracted by *Sox1*-3 activity. Nat Neurosci. 2003;6:1162–1168. doi: 10.1038/nn1131.

11 Graham V, Khudyakov J, Ellis P, Pevny L. SOX2 functions to maintain neural progenitor identity. Neuron. 2003;39:749–765. doi: 10.1016/S0896-6273(03)00497-5.

12 Taranova OV, *et al*. SOX2 is a dose-dependent regulator of retinal neural progenitor competence. Genes Dev. 2006;20:1187–1202. doi: 10.1101/gad.1407906.

13 Niu W, *et al*. In vivo reprogramming of astrocytes to neuroblasts in the adult brain. Nat Cell Biol. 2013;15:1164–1175. doi: 10.1038/ncb2843.

14 Oosterveen T, *et al*. SoxB1-driven transcriptional network underlies neural-specific interpretation of morphogen signals. Proc Natl Acad Sci USA. 2013;110:7330–7335. doi: 10.1073/pnas.1220010110.

15 Su Z, Niu W, Liu ML, Zou Y, Zhang CL. In vivo conversion of astrocytes to neurons in the injured adult spinal cord. Nat Commun. 2014;5:3338.

16 Cai N, Li M, Qu J, Liu GH, Izpisua Belmonte JC. Post-translational modulation of pluripotency. J Mol Cell Biol. 2012;4:262–265. doi: 10.1093/jmcb/mjs031.

17 Uchikawa M, Ishida Y, Takemoto T, Kamachi Y, Kondoh H. Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. Dev Cell. 2003;4:509–519. doi: 10.1016/S1534-5807(03)00088-1.

18 Catena R, *et al*. Conserved POU binding DNA sites in the Sox2 upstream enhancer regulate gene expression in embryonic and neural stem cells. J Biol Chem. 2004;279:41846–41857. doi: 10.1074/jbc.M405514200.

19 Okamoto R, Uchikawa M, Kondoh H. Sixteen additional enhancers associated with the chicken Sox2 locus outside the central 50-kb region. Dev Growth Differ. 2015;57:24–39. doi: 10.1111/dgd.12185.

20 Zhou HY, *et al*. A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. Genes Dev. 2014;28:2699–2711. doi: 10.1101/gad.248526.114.

21 Avilion AA, *et al*. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev. 2003;17:126–140. doi: 10.1101/gad.224503.

22 Ellis P, *et al*. SOX2, a persistent marker for multipotential neural stem cells derived from embryonic stem cells, the embryo or the adult. Dev Neurosci. 2004;26:148–165. doi: 10.1159/000082134.

23 Kang W, Hebert JM. A Sox2 BAC transgenic approach for targeting adult neural stem cells. PLoS One. 2012;7:e49038. doi: 10.1371/journal.pone.0049038.

24 van Leeuwen S, Mikkers H. Long non-coding RNAs: Guardians of development. Differentiation. 2010;80:175–183. doi: 10.1016/j.diff.2010.07.003.

25 Brockdorff N. Noncoding RNA and Polycomb recruitment. RNA. 2013;19:429–442. doi: 10.1261/rna.037598.112.

26 Wang KC, *et al*. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature. 2011;472:120–124. doi: 10.1038/nature09819.

27 Latos PA, *et al*. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science. 2012;338:1469–1472. doi: 10.1126/science.1228110.

28 Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature. 2007;445:666–670. doi: 10.1038/nature05519.

29 Crevillen P, Sonmez C, Wu Z, Dean C. A gene loop containing the floral repressor FLC is disrupted in the early phase of vernalization. EMBO J. 2013;32:140–148. doi: 10.1038/emboj.2012.324.

30 Fantes J, *et al*. Mutations in SOX2 cause anophthalmia. Nat Genet. 2003;33:461–463. doi: 10.1038/ng1120.

31 Amaral PP, *et al*. Complex architecture and regulated expression of the *Sox2ot* locus during vertebrate development. RNA. 2009;15:2013–2027. doi: 10.1261/rna.1705309.

32 Askarian-Amiri ME, *et al*. Emerging role of long non-coding RNA SOX2OT in SOX2 regulation in breast cancer. PLoS One. 2014;9:e102140. doi: 10.1371/journal.pone.0102140.

33 Hou Z, *et al*. A long noncoding RNA *Sox2ot* regulates lung cancer cell proliferation and is a prognostic indicator of poor survival. Int J Biochem Cell Biol. 2014;53:380–388. doi: 10.1016/j.biocel.2014.06.004.

34 Shahryari A, *et al*. Two novel splice variants of SOX2OT, SOX2OT-S1, and SOX2OT-S2 are coupregulated with SOX2 and OCT4 in esophageal squamous cell carcinoma. Stem Cells. 2014;32:126–134. doi: 10.1002/stem.1542.

35 Tosetti V, *et al*. Transcriptional role of androgen receptor in the expression of long non-coding RNA Sox2OT in neurogenesis. PLoS One. 2017;12:e0180579. doi: 10.1371/journal.pone.0180579.

36 Hanada K, *et al*. sORF finder: a program package to identify small open reading frames with high coding potential. Bioinformatics. 2010;26:399–400. doi: 10.1093/bioinformatics/btp688.

37 Smith JE, *et al*. Translation of small open reading frames within unannotated RNA transcripts in Saccharomyces cerevisiae. Cell Rep. 2014;7:1858–1866. doi: 10.1016/j.celrep.2014.05.023.

38 Woolfe A, *et al*. Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol. 2005;3:e7. doi: 10.1371/journal.pbio.0030007.

39 Carninci P, Hayashizaki Y. Noncoding RNA transcription beyond annotated genes. Curr Opin Genet Dev. 2007;17:139–144. doi: 10.1016/j.gde.2007.02.008.

40 Marques AC, Ponting CP. Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness. Genome Biol. 2009;10:R124. doi: 10.1186/gb-2009-10-11-r124.

41 Chodroff RA, *et al*. Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. Genome Biol. 2010;11:R72. doi: 10.1186/gb-2010-11-7-r72.

42 Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods. 2008;5:877–879. doi: 10.1038/nmeth.1253.

43 Wood HB, Episkopou V. Comparative expression of the mouse *Sox1*, Sox2 and *Sox3* genes from pre-gastrulation to early somite stages. Mech Dev. 1999;86:197–201. doi: 10.1016/S0925-4773(99)00116-1.

44 Duinsbergen D, Eriksson M, t Hoen PA, Frisen J, Mikkers H. Induced pluripotency with endogenous and inducible genes. Exp Cell Res. 2008;314:3255–3263. doi: 10.1016/j.yexcr.2008.06.024.

45 Silva J, Chambers I, Pollard S, Smith A. Nanog promotes transfer of pluripotency after cell fusion. Nature. 2006;441:997–1001. doi: 10.1038/nature04914.

46 West JA, *et al*. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. Mol Cell. 2014;55:791–802. doi: 10.1016/j.molcel.2014.07.012.

47 Rinn JL, Chang HY. Genome Regulation by Long Noncoding RNAs. Annu Rev Biochem. 2012;81:145–166. doi: 10.1146/annurev-biochem-051410-092902.

48 Bernstein BE, *et al*. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006;125:315–326. doi: 10.1016/j.cell.2006.02.041.

49 Voigt P, Tee WW, Reinberg D. A double take on bivalent promoters. Genes Dev. 2013;27:1318–1338. doi: 10.1101/gad.219626.113.

50 Dean A. On a chromosome far, far away: LCRs and gene expression. Trends Genet. 2006;22:38–45. doi: 10.1016/j.tig.2005.11.001.
51 Zhang Y, *et al*. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature. 2013;504:306–310. doi: 10.1038/nature12716.
52 Li Y, *et al*. CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. PLoS One. 2014;9:e114485. doi: 10.1371/journal.pone.0114485.
53 Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. Cell. 2013;152:1237–1251. doi: 10.1016/j.cell.2013.02.014.
54 Maggio I, Goncalves MA. Genome editing at the crossroads of delivery, specificity, and fidelity. Trends Biotechnol. 2015;33:280–291. doi: 10.1016/j.tibtech.2015.02.011.
55 Hagey DW, Muhr J. Sox2 acts in a dose-dependent fashion to regulate proliferation of cortical progenitors. Cell Rep. 2014;9:1908–1920. doi: 10.1016/j.celrep.2014.11.013.
56 Ying QL, *et al*. The ground state of embryonic stem cell self-renewal. Nature. 2008;453:519–523. doi: 10.1038/nature06968.
57 Messemaker TC, *et al*. A novel long non-coding RNA in the rheumatoid arthritis risk locus TRAF1-C5 influences C5 mRNA levels. Genes Immun. 2016;17:85–92. doi: 10.1038/gene.2015.54.
58 Bain G, Kitchens D, Yao M, Huettner JE, Gottlieb DI. Embryonic stem cells express neuronal properties in vitro. Dev Biol. 1995;168:342–357. doi: 10.1006/dbio.1995.1085.
59 Verma I, Rashid Z, Sikdar SK, Seshagiri PB. Efficient neural differentiation of mouse pluripotent stem cells in a serum-free medium and development of a novel strategy for enrichment of neural cells. Int J Dev Neurosci. 2017;61:112–124. doi: 10.1016/j.ijdevneu.2017.06.009.
60 Engberg N, Kahn M, Petersen DR, Hansson M, Serup P. Retinoic acid synthesis promotes development of neural progenitors from mouse embryonic stem cells by suppressing endogenous, Wnt-dependent nodal signaling. Stem Cells. 2010;28:1498–1509. doi: 10.1002/stem.479.
61 Conti L, *et al*. Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. PLoS Biol. 2005;3:e283. doi: 10.1371/journal.pbio.0030283.
62 Semrau S, *et al*. FuseFISH: robust detection of transcribed gene fusions in single cells. Cell Rep. 2014;6:18–23. doi: 10.1016/j.celrep.2013.12.002.
63 Lee TI, Johnstone SE, Young RA. Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat Protoc. 2006;1:729–748. doi: 10.1038/nprot.2006.98.
64 Stadhouders R, *et al*. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. Nat Protoc. 2013;8:509–524. doi: 10.1038/nprot.2013.018.
65 Denholtz M, *et al*. Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. Cell Stem Cell. 2013;13:602–616. doi: 10.1016/j.stem.2013.08.013.
66 Werner MS, Ruthenburg AJ. Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. Cell Rep. 2015;12:1089–1098. doi: 10.1016/j.celrep.2015.07.033.
67 Mikkelsen TS, *et al*. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007;448:553–560. doi: 10.1038/nature06008.

Chapter 8
Summarizing discussion

Numerous studies have contributed to our current understanding of autoimmune diseases (AIDs), however, pathogenesis of many AIDs can still not be fully explained. Both genetic factors and environmental factors are involved in the onset of autoimmunity. Which mechanisms explain the contribution of these genetic and environmental factors to disease pathogenesis, and how the different factors interplay remain unanswered key questions. The studies presented in this thesis aimed at identifying and unravelling some of the enigmatic mechanisms in rheumatoid arthritis (RA) and systemic sclerosis (SSc).

Epigenetic changes are thought to play a role in passing on environmental influences to gene expression alterations that can contribute to disease. In **Chapter 2** of this thesis, we investigated whether monocytes from diagnosed but yet untreated RA patients contain a distinct, disease-related, epigenetic signature of genes associated with RA. No large epigenetic differences were observed between RA and healthy monocytes indicating that such differences are either small or not present in the tested cell type for TNFα and IL6. However, epigenetic differences were observed in RA patients in other cell types indicating that epigenetic changes can play an important role in RA. DNA hypomethylation was found in synovial fibroblast from RA patients and indicate that cells in a disease-affected environment may display epigenetic differences[1]. Therefore, also monocytes or other cell types in the synovium may display differences and thereby contribute to disease pathogenesis. It would be interesting to investigate whether epigenetic differences are also present and maintained in precursor cells (like CD34+ cells). Upon differentiation, these cells may end up in the joints and trigger or enhance the inflammatory status found in RA patients[2,3]. Which cell types do contain these epigenetic traces, how they obtain these marks and how we can restore an autoimmune epigenetic landscape are topics for future studies. Moreover, another key question is whether these epigenetic marks are present prior to the onset of the disease or whether these marks are a consequence of disease pathogenesis. This question may be answered by longitudinal retrospective studies in which the responsible cell types have been collected. In case these marks are present prior to the onset of a disease, they may also play a crucial role in the diagnoses and treatment of autoimmunity by opening early treatment options. Together, further efforts investigating how and which epigenetic changes are involved in disease pathogenesis on a genome-wide level and a cell-type specific manner are needed to increase our

understanding in disease pathogenesis and may reveal early diagnostic markers or open up novel treatment options.

Large genetic studies containing the genetic information of over 100.000 individuals have been performed to relate variants and genes to a role in disease pathogenesis of rheumatoid arthritis[4]. These genetic population studies can identify hundreds of variants in a single locus that all associate with disease due to high linkage disequilibrium. Identifying the causal SNPs is often difficult as the highest associated variant (lead SNP) of a disease associated locus is not necessarily the causing variant[5]. Revealing which functional mechanisms shelter behind associated SNPs aids in understanding how genes are affected and which pathways may play a role in disease pathogenesis. **Chapter 3** of this thesis reviews identified variants contributing functionally to disease, and the involved pathways that are hypothesized to play a role in RA. For example, the coding variant (Arg620Trp) in PTPN22 was shown to affect both BCR and TCR signalling[6]. Moreover, several variants in different genes have shown to affect NF-kB signalling, including: the variants Val194Ala and Pro175Leu in NFKBIE, variant Phe127Cys in TNFA3 and variant Ala288Thr in RTKN2. Similar evidence for the involvement of these pathways came forward from gene enrichment analysis of candidate genes located in the 100 associated risk loci which identified T-cell receptor (TCR) signalling, NF-κb signalling and JAK-STAT signalling as the most enriched processes (**Chapter 3**)[6]. Several other studies have investigated the role of these pathways in context of autoimmunity[7,8]. In the JAK-STAT signalling cascade, STAT is phosphorylated by JAK proteins resulting in the activation of pro-inflammatory cytokines thereby promoting the inflammatory state in RA patients[9]. Inhibitors of this cascade have with success been tested as therapeutics reducing the level of pro inflammatory cytokines[10]. Tofacitinib, a JAK-STAT inhibitor has received FDA approval and several other inhibitors are being tested in clinical trials[11–13]. Similarly, functional studies have highlighted enhanced NF-κB activity and defective TCR signalling in RA patients[1415]. Studies are undergoing investigating potential therapeutics targeting both TCR receptor signalling and the NF-κB cascade[16–19]. Together, we hypothesize that non-HLA RA-associated variants in these genes and pathway are responsible for a decreased immune activation threshold and for disturbing a healthy ratio between pro and anti-inflammatory cytokines increasing the probability of developing RA. Although for some RA-associated variants the casual mechanisms has been revealed, future

studies should be conducted for the remaining variants. Thereby, understanding the influence of variants in these genes from the identified pathways might also explain why some of the used therapeutics is not beneficial for all RA patients and stimulates the research into personal medicine within the field of RA and autoimmunity.

However for the majority of risk loci, the causal mechanisms for their association with RA remain elucidative. One of these loci is the *TRAF1-C5* locus which contains multiple RA-associating variants in high linkage disequilibrium of which the causal variant has not yet been identified. Although variants in *C5* have been identified as variants affecting C5 function, these variants do not significantly associate with RA. It is therefore unlikely that these variants can explain the association of this region with RA as described in **Chapter 4**. The *TRAF1-C5* locus lacks RA-associated variants that change amino acids of the nearby candidate genes and therefore no functional mechanism have been identified. In **Chapter 5** of this thesis we describe our discovery of a novel gene named *C5T1lncRNA* in this region. Interestingly, two SNPs are located in the RNA sequence of this presumably long non-coding RNA (lncRNA). Non-coding RNAs do not translate into proteins and these SNPs are therefore not identified as amino acid changing variants. Nonetheless, SNPs in lncRNAs can be functional variants as several studies have shown that SNPs can alter i.) the binding potential of the lncRNA, ii.) the structure of the lncRNA and iii.) lncRNA expression levels[20–22]. Moreover, the identified lncRNA in the *TRAF1-C5* locus was found to be expressed and functional in RA-relevant cell types as synovial fibroblasts and may therefore have a functional role in RA pathogenesis. We speculate a mechanism in which variants in *C5T1-lncRNA* might interfere with the function of this gene. In **Chapter 5**, we found that decreasing levels of C5T1lncRNA also decreased levels of the nearby gene *C5* indicating a regulatory role. Variants in *C5T1lncRNA* might therefore interfere with this regulatory role and might thus also affect the function of C5, a potent pro-inflammatory immune gene. Future studies should be designed to investigate the effect of the variants in the *TRAF1-C5* locus and what consequences this brings for *C5*. Thereby, we cannot rule out the possibility that variants in the *TRAF1-C5* region influences either with *C5* and *TRAF1* via other mechanisms. Several eQTL effects were found from variants in the *TRAF1-C5* region[23,24]. These variants could interfere with C5 and TRAF1 levels by for example influencing the mRNA stability or by interfering with transcription

factors binding sites. Such mechanisms could function as causal mechanisms for RA independent of *C5T1lncRNA*. Additionally, a cell-type specific manner in which variants affect genes in the *TRAF1-C5* locus is possible[25]. *C5T1lncRNA* is highly expressed in the liver, similar to *C5*, but *C5T1lncRNA* is also strongly induced by LPS in monocytes, similar to *TRAF1*, illustrating the complex nature of this locus[26]. In order to aid in addressing the functional mechanisms of such loci, large studies have been set up to collect cell type specific expression in hundreds of cell types[27]. Currently, FANTOM5, TiGER and GTEX are large databases that provide such expression data of over 20.000 genes in more than 400 cell types and over 100 different tissues providing useful platforms for future expression studies[27–29].

To identify functional variants originating from genome wide association studies (GWAS) and to understand genomic variation, large studies have been set up focussing on gene expression changes linked to genomic variation, also known as eQTL studies. A large study that included over 5000 individuals identified that genetic variations can influence gene expression of genes, both in *cis* and in *trans*[30]. Another large study investigated expression changes specifically in monocytes from over 1000 individuals and reported similar findings[31]. These studies provide a useful platform and starting point for the unravelling of functional genetic variants. Such studies also provide insight into which cell types play a role in disease by investigating cell-type specific eQTLs. A recent study investigated cell type specific eQTLs in monocytes and B-cells and showed that disease associating variants can have functional consequences in a cell type specific manner[32]. Moreover, Raj *et al*. investigated cell type specific traits in T-cells and monocytes and identified that many variants associated with RA specifically influenced the expression of genes in T-cells[33]. From these studies it has been concluded that variants often display cell specific traits and may indicate which cell types play a role in disease pathogenesis. Additional genetic evidence showed that T-cells play an important role in RA. Overlapping disease-associating variants with the presence of active or repressing histone modifications in a cell type specific manner provides indications in which cell type, which variants are being accessible. Farh *et al*. found that RA-associating variants display histone modifications that are enriched in T-cells, B-cells and lymphoblastoid cells in a comparison with 33 different cell-types[34]. Finally, examining lncRNA expression in RA-associated loci has been linked to T-cells as Hrdlicknova *et al*. has shown that the lncRNAs located in associated regions are

often specifically expressed in T-cells[35]. These studies illustrate which cell types may be responsible and indicate that not only coding genes but also non-coding genes are potential disease genes when affected by variants. Although enrichment statistics and gene coexpression are not conclusive with regard to causality and functionality, additional functional studies are necessary. Nonetheless, it is likely that development to RA is affected by defects in multiple cell types of which T cells and T-cell activation play an important and determining role. Genetic variants likely affect genes in a cell specific manner resulting together with other cellular defects and environmental alterations in an increased susceptibility to RA.

Aside from genetic studies, RNA sequencing of disease-relevant tissues can also highlight genes and pathways involved in disease pathogenesis. In **Chapter 6,** the RNA of skin from SSc patients was compared with skin from healthy donors, and resulted in the identification of both deregulated coding and non-coding genes. In this chapter specifically non-coding genes were investigated and hundreds of deregulated lncRNAs were observed. Among these, several lncRNAs were validated using a replication dataset, including AGAP2-AS1, CTBP1-AS2 and OTUD6B-AS1. These genes are classified as antisense genes and in both studies, also their sense gene was deregulated. Although no functional assessment was performed in this study, we hypothesize that such deregulated gene pairs play a role in the disease pathogenesis of SSc. In such a model the deregulated antisense gene fails to maintain its regulatory role on its opposing sense gene resulting in a deregulated gene pair leading to depending on its function to disease pathogenesis. Coinciding with this model is the high correlation that was found between the expression of both genes within such gene pairs in our study and other studies[36–38]. Overall, we hypothesize that some of these lncRNAs either are involved with functions contributing to SSc directly, or by influencing other coding genes thereby contributing to SSc pathogenesis. Although thousands of long non-coding RNAs have been discovered, very few molecular mechanisms have yet been identified. lncRNAs can have a diverse set of functions and interfere not only in disease pathogenesis but also developmental processes. Like described in **Chapter 7**, *Sox2ot*, a lncRNA overlapping *Sox2*, interferes with *Sox2* gene transcription. Sox2ot is a gene that is located near enhancer and transcription regions that are important for Sox2 expression. Expression of Sox2ot is hypothesized to interfere with the transcriptional process of Sox2

thereby regulating its levels. In a developmental point of view, similar mechanisms are possible for other development genes. For example, Sox1 and Sox4 display a similar genetic landscape and might therefore also be under regulation of non-coding RNA transcripts. The hypothesized mechanism of Sox2ot that came forward from the study in **Chapter 7** was interference of enhancer regions by altering DNA-looping events. Currently, studies are on-going to reveal in-depth genetic landscapes and cross-communication of genes, enhancers, transcription factors, via chromatin-loops[39,40]. Novel methods allow more detailed overview of this genetic landscape and will aid in unravelling non-coding RNA functions and disease mechanisms. Together, our studies contribute to a better understanding of how genes are regulated, which DNA regions are responsible for gene activation and gene silencing and whether non-coding genes might be involved.

Unravelling the function of lncRNAs is essential to understand their role and involvement in development but also in diseases like autoimmunity. Currently several laboratories have set up large scale experiments to investigate these functions, especially in cancer by evaluating lncRNAs involved in cell growth[41,42]. These studies have identified numerous lncRNAs functionally involved in cell growth in several cancer cell lines. However not all lncRNAs function through interference with cell growth and therefore similar studies should be set up focusing on other cellular functions. An example would be to knock down levels of (or knockout) lncRNAs in immune cell types followed by various immune activation signals to identify which lncRNAs are involved in the immune response. In the near future, such studies will be performed and will be aided by the revolution of CRISPR technology allowing largescale knockdown technology.

More and more lncRNAs are being identified as deregulated genes in disease and development which opens the possibility to use them as diagnostic markers or therapeutic targets. Although, non-coding genes are overall lower expressed compared to coding genes, they also possess characteristics that will prefer non-coding genes over coding genes as future drug targets. For example their cell-type specificity allows drugs to be effective in one cell-type only, preventing unwanted side effects in other cell types or tissues. Especially in cancers, where cancer-specific lncRNA expression can be used as a therapeutic targets thereby leaving healthy tissue unaffected. The first report has already shown that targeting a lncRNA known as MALAT by antisense oligo nucleotides was able to

prevent lung cancer metastasis in mice displaying the feasibility of targeting lncRNAs[43]. Other potential intervention approaches through lncRNAs that are in pre-clinical development include siRNAs, aptamers, ribozymes or small molecules and are reviewed in ref[44]. As lncRNAs are often highly expressed in specific diseased cells (like cancer cells) they can also be used as biomarkers and for diagnostic purposes. A diagnostic test using an overexpressed lncRNA is currently under development and is applicable for the diagnoses of prostate cancer[45]. This test can measure levels of PCA3, a prostate specific lncRNA overexpressed in prostate cancer, in the urine of patients[45]. With rapidly advancing technology it will be easier to detect and target lncRNAs and therefore an increasing amount of specific biomarkers for early diagnoses, better prognostic prediction and more efficient therapy will undoubtedly be available in future clinical applications.

The studies presented in this thesis contributed to the identification of lncRNAs involved in disease pathogenesis. Although non-coding RNAs are overall lower expressed, still they may regulate crucial functions and should not be disregarded merely based on present abundances. Future single-cell sequencing studies will be able to gather detailed information regarding non-coding RNAs and their mechanisms in cell specific manners. Together the reducing costs for sequencing, the increasing single cell resolution to study gene expression and the efficient single cell isolation technology provide a highly accurate platform to study both basic and translational research. Expression profiles of both coding and non-coding RNAs on single cell levels may aid in the identification and characterisation of novel and existing cell types. Therefore further unravelling mechanisms by which non-coding RNAs function not only lead to insight in disease development but we hypothesise the idea that non-coding genes will one day be used as target genes in future therapies, including diseases of autoimmunological nature. Finally, if epigenetic alterations (such as histone modifications or non-coding RNA dysregulation) occur years before the onset of a disease, they may be better therapeutic targets prevent the disease compared to current medicines who are often used to supress the disease or to treat the symptoms only.

References

1 Karouzakis E, Gay RE, Michel B a, Gay S, Neidhart M. DNA hypomethylation in rheumatoid arthritis synovial fibroblasts. *Arthritis Rheum* 2009; **60**: 3613–22.

2 Hirohata S, Yanagida T, Nampei A, Kunugiza Y, Hashimoto H, Tomita T *et al.* Enhanced generation of endothelial cells from CD34+ cells of the bone marrow in rheumatoid arthritis: Possible role in synovial neovascularization. *Arthritis Rheum* 2004; **50**: 3888–3896.

3 Hirohata S, Yanagida T, Tomita T, Ochi T, Nakamura H, Yoshino S. Bone marrow CD34+ progenitor cells from rheumatoid arthritis patients give rise to spontaneous transformation of peripheral blood B cells from Epstein-Barr virus infected healthy individuals. *Arthritis Rheum* 1996; **39**: 1002.

4 Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2013. doi:10.1038/nature12873.

5 Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome TL - 22. *Genome Res* 2012; **22 VN**-**r**: 1748–1759.

6 Messemaker TC, Huizinga TW, Kurreeman F. Immunogenetics of rheumatoid arthritis: Understanding functional implications. *J Autoimmun* 2015; **64**: 7–14.

7 Banerjee S, Biehl A, Gadina M, Hasni S, Schwartz DM. JAK–STAT Signaling as a Target for Inflammatory and Autoimmune Diseases: Current and Future Prospects. *Drugs* 2017; **77**: 521–546.

8 Makarov SS. NF-kappa B in rheumatoid arthritis: a pivotal regulator of inflammation, hyperplasia, and tissue destruction. *Arthritis Res* 2001; **3**: 200–6.

9 Darnell J, Kerr I, Stark G. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science (80-)* 1994; **264**: 1415–1421.

10 Malemud CJ. Inhibitors of JAK for the treatment of rheumatoid arthritis: rationale and clinical data. *Clin Investig (Lond)* 2012; **2**: 39–47.

11 Kremer JM, Emery P, Camp HS, Friedman A, Wang L, Othman AA *et al.* A Phase IIb Study of ABT-494, a Selective JAK-1 Inhibitor, in Patients With Rheumatoid Arthritis and an Inadequate Response to Anti–Tumor Necrosis Factor Therapy. Arthritis Rheumatol. 2016; **68**: 2867–2877.

12 Boyle DL, Soma K, Hodge J, Kavanaugh A, Mandel D, Mease P *et al.* The JAK inhibitor tofacitinib suppresses synovial JAK1-STAT signalling in rheumatoid arthritis. *Ann Rheum Dis* 2015; **74**: 1311–1316.

13 Vanhoutte F, Mazur M, Voloshyn O, Stanislavchuk M, Van der Aa A, Namour F *et al.* Efficacy, safety, pharmacokinetics, and pharmacodynamics of filgotinib, a selective Janus kinase 1 inhibitor, after short-term treatment of rheumatoid arthritis: Results of two randomized Phase IIA trials. *Arthritis Rheumatol* 2017; **69**: 1949–1959.

14 Maurice MM, Lankester a C, Bezemer a C, Geertsma MF, Tak PP, Breedveld FC *et al.* Defective TCR-mediated signaling in synovial T cells in rheumatoid arthritis. *J Immunol* 1997; **159**: 2973–8.

15 Lawrence T, Gilroy DW, Colville-Nash PR, Willoughby DA. Possible new role for NF-kappaB in the resolution of inflammation. Nat. Med. 2001; **7**: 1291–1297.

16 Esensten JH, Helou YA, Chopra G, Weiss A, Bluestone JA. CD28 Costimulation: From Mechanism to Therapy. *Immunity* 2016; **44**: 973–988.

17 Thomas R, Turner M, Cope AP. High avidity autoreactive T cells with a low signalling capacity through the T-cell receptor: central to rheumatoid arthritis pathogenesis? *Arthritis Res Ther* 2008; **10**: 210.

18 Zhuang Y, Liu J, Ma P, Bai J, Ding Y, Yang H *et al.* Tamarixinin a alleviates joint destruction of rheumatoid arthritis by blockade of MAPK and NF-κB activation. Front. Pharmacol. 2017; **8**. doi:10.3389/fphar.2017.00538.

19 Okamoto H, Yoshio T, Kaneko H, Yamanaka H. Inhibition of NF-KappaB signaling by fasudil as a potential therapeutic strategy for rheumatoid arthritis. *Arthritis Rheum* 2010; **62**: 82–92.

20 Mirza AH, Kaur S, Brorsson CA, Pociot F. Effects of GWAS-associated genetic variants on lncRNAs within IBD and T1D candidate loci. *PLoS One* 2014; **9**. doi:10.1371/journal.pone.0105723.

21 Almlöf JC, Lundmark P, Lundmark A, Ge B, Pastinen T, Goodall AH *et al.* Single nucleotide polymorphisms with cis-regulatory effects on long non-coding transcripts in human primary monocytes. *PLoS One* 2014; **9**: e102612.

22 Bhartiya D, Scaria V. Genomics Genomic variations in non-coding RNAs : Structure , function and regulation. *Genomics* 2016; **107**: 59–68.

23 van Steenbergen HW, Rodríguez-Rodríguez L, Berglin E, Zhernakova A, Knevel R, Ivorra-Cortés J *et al.* A genetic study on C5-TRAF1 and progression of joint damage in rheumatoid arthritis. *Arthritis Res Ther* 2015; **17**: 1.

24 Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* 2016; **48**: 481–487.

25 Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet* 2012; **44**: 502–510.

26 Messemaker TC, Frank-Bertoncelj M, Marques RB, Adriaans a, Bakker a M, Daha N *et al.* A novel long non-coding RNA in the rheumatoid arthritis risk locus TRAF1-C5 influences C5 mRNA levels. *Genes Immun* 2016; **17**: 85–92.

27 Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S *et al.* Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol* 2015; **16**: 22.

28 Liu X, Yu X, Zack DJ, Zhu H, Qian J. TiGER: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics* 2008; **9**: 271.

29 Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013; **45**: 580–585.

30 Westra H, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013; **45**: 1238–1243.

31 Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R *et al.* Genetics and beyond - the transcriptome of human monocytes and disease susceptibility. *PLoS One* 2010; **5**. doi:10.1371/journal.pone.0010693.

32 Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A *et al.* Genetics of gene expression in primary immune cells identifies cell type – specific master regulators and roles of HLA alleles. *Nat Genet* 2012; **44**: 502–510.

33 Raj T, Rothamel K, Mostafavi S, Ye C, Lee MN, Replogle JM *et al.* Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science (80-)* 2014; **344**: 519–23.

34 Farh KK-H, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 2015; **518**: 337–43.

35 Hrdlickova B, Kumar V, Kanduri K, Zhernakova D V, Tripathi S, Karjalainen J *et al.* Expression profiles of long non-coding RNAs located in autoimmune disease-associated regions reveal immune cell-type specificity. *Genome Med* 2014; **6**: 88.

36 Goyal A, Fiškin E, Gutschner T, Polycarpou-Schwarz M, Groß M, Neugebauer J *et al.* A cautionary tale of sense-antisense gene pairs: independent regulation despite inverse correlation of expression. *Nucleic Acids Res* 2017; **45**: 12496–12508.

37 Messemaker TC, Chadli L, Cai G, Goelela VS, Boonstra M, Dorjée AL *et al.* Antisense long non-coding RNAs are deregulated in skin tissue of patients with systemic sclerosis. *J Invest Dermatol* 2017. doi:10.1016/j.jid.2017.09.053.

38 Balbin OA, Malik R, Dhanasekaran SM, Prensner JR, Cao X, Wu Y *et al.* The landscape of antisense gene expression in human cancers. *Genome Res* 2015; **25**: 1068–1079.

39 Zhang H, Li F, Jia Y, Xu B, Zhang Y, Li X *et al.* Characteristic arrangement of nucleosomes is predictive of chromatin interactions at kilobase resolution. *Nucleic Acids Res* 2017; : 1–13.

40 Barutcu AR, Fritz AJ, Zaidi SK, van Wijnen AJ, Lian JB, Stein JL, Nickerson JA, Imbalzano AN SG. C-ing the Genome: A Compendium of Chromosome Conformation Capture Methods to Study Higher-Order Chromatin Organization. *J Cell Physiol* 2016; **21**: 133–136.

41 Zhu S, Li W, Liu J, Chen C, Liao Q, Xu P *et al.* Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library. *Nat Biotechnol* 2016; **34**: 1279–1286.

42 Liu SJ, Liu SJ, Horlbeck MA, Cho SW, Birk HS, Malatesta M *et al.* CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science (80-)* 2016; **7111**: aah7111.

43 Gutschner T, Hämmerle M, Eißmann M, Hsu J, Kim Y, Hung G *et al.* The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res* 2013; **73**: 1180–1189.

44 Parasramka MA, Maji S, Matsuda A, Yan IK, Patel T. Pharmacology & Therapeutics Long non-coding RNAs as novel targets for therapy in hepatocellular carcinoma. *Pharmacol Ther* 2016; **161**: 67–78.

45 Lee GL, Dobi A, Srivastava S. Diagnostic performance of the PCA3 urine test. *Prostate Cancer* 2011; **8**: 123–124.

Chapter 9
Addendum:

Nederlandse samenvatting
List of publications
Curriculum Vitae
Dankwoord

Nederlandse samenvatting

Auto-immuunziekten staan bekend als ziekten waarbij het immuunsysteem cellen of stoffen van ons eigen lichaam herkent als lichaamsvreemd en daarbij dus het eigen lichaam aanvalt. Auto-immuunziekten zijn veel voorkomend en kunnen een grote impact hebben op het dagelijkse leven van een patiënt. Ondanks dat er veel verschillende medicijnen zijn om het immuunsysteem onder controle te houden, is het genezen van deze ziekten nog niet mogelijk. Om auto-immuunziekten te voorkomen en/of genezen is het begrijpen van het ziekteproces cruciaal. In de afgelopen 60 jaar is door onderzoek al veel duidelijk geworden over de ontwikkeling van auto-immuunziekten. Zo zijn er verschillende omgevingsfactoren geïdentificeerd die een rol spelen bij het ontstaan van auto-immuunziekten. Meerdere studies hebben laten zien dat roken een verhoogde kans geeft op het ontwikkelen van bijvoorbeeld reumatoïde artritis en blootstelling aan kwarts (ook wel siliciumdioxide genoemd) wordt geassocieerd met het ontwikkelen van systemische sclerose. Daarnaast wordt ook gedacht dat de dagelijkse interactie tussen het immuunsysteem en pathogenen (zoals bacteriën en virussen) een rol speelt bij auto-immuunziekten en het immuunsysteem uit balans kan halen. Tot slot weten we ook dat een genetische aanleg de kans op het ontstaan van auto-immuunziekten verhoogt. Deze ontdekkingen hebben bijgedragen aan onze kennis over auto-immuunziekten en heeft direct geleid tot nieuwe, succesvolle medicijnen die momenteel gebruikt worden in de kliniek om deze ziekten te behandelen. Echter meer onderzoek is nodig om de auto-immuunziekten nog beter te begrijpen zodat we in de toekomst deze aandoeningen volledig kunnen voorkomen en genezen. In het in dit proefschrift beschreven onderzoek staat het ontdekken en begrijpen van factoren die een rol kunnen spelen in auto-immuunaandoeningen reumatoïde artritis en systemische sclerosis centraal.

Uit voorgaand onderzoek is gebleken dat hoge niveaus van de factoren TNFα en IL6 in het bloed van reumapatiënten circuleren, factoren die bijdragen aan de activatie van het immuunsysteem. In **hoofdstuk 2** is de hypothese onderzocht of reumapatiënten deze hoge levels TNFα en IL6 hebben door een afwijkende regulatie van deze genen. Epigenetica speelt een belangrijke rol bij de regulatie van genen en staat bekend als de omkeerbare erfelijke veranderingen die niet de DNA-sequentie zelf beïnvloedt. Deze epigenetische veranderingen kunnen een

lange tijd aanwezig blijven op bepaalde genen en daardoor bijdragen aan consistent hogere levels van deze genen. Om te onderzoeken of dit het geval is bij reumapatiënten, zijn epigenetische veranderingen van de genen *TNFα* en *IL6* onderzocht in cellen van reuma patiënten en gezonde controles. Omdat monocyten bekend staan als producenten van TNFα en IL6 is dit celtype onderzocht in beide groepen. Hieruit blijkt echter dat deze monocyten afkomstig uit het bloed geen verschillen laten zien op het gebied van gen regulatie en evenredige niveaus van deze epigenetische modificaties aanwezig waren. Dit resultaat werd verder bevestigd door te laten zien dat monocyten van de geteste reuma patiënten en gezonde mensen ook na stimulatie vergelijkbare niveaus van TNFα en IL6 lieten zien.

Naast epigenetische veranderingen, zijn in dit proefschrift ook veranderingen in het DNA (genetisch) die een risico op het ontwikkelen van reuma geven onderzocht. **Hoofdstuk 3** geeft een overzicht van de literatuur van de genetische bijdragen aan reuma. Samengevat zien we dat genen die in genetische geassocieerde gebieden liggen vooral een rol spelen in immunologische gerelateerde processen zoals T-cel signalering, JAK-STAT signalering en NF-κB signalering. Verder worden in dit hoofdstuk mogelijk nieuwe bijdragende genen geïntroduceerd, ook wel bekend als niet-coderende genen. Recentelijk is ontdekt dat de meerderheid van de genen in ons genoom wel wordt afgeschreven naar RNA maar niet vertaald word naar een functioneel eiwit. Deze niet-coderende RNA moleculen worden belangrijke cellulaire functies toegedicht maar zijn tot nu toe vaak buiten beschouwing gebleven.

In **hoofdstuk 4 en 5** is een specifiek genetisch risicogebied voor reuma die zowel de genen *TRAF1* als ook *C5* bevat verder onderzocht. Genetische varianten in dit gebied (*TRAF1-C5* locus) treffen alleen niet-coderende nucleïnezuren waardoor de bijdrage van dit genetisch gebied aan reuma vooralsnog speculatief is. In **hoofdstuk 5** wordt een nieuw niet-coderend gen in dit gebied beschreven en is onderzocht hoe dit gen bij zou kunnen dragen aan reuma. Dit gen (genaamd *C5T1-lcnRNA*) wordt met name afgeschreven in lever- en immuuncellen en de niveaus van dit gen gaan omhoog na immuun-activatie. De functie van *C5T1-lcnRNA* lijkt gerelateerd aan regulatie van het nabij gelegen gen genaamd *C5* welke een prominente rol heeft in het afweersysteem. Twee genetische varianten die associëren met reuma liggen in *C5T1lncRNA* en kunnen dus

mogelijk zo de functie van *C5T1lncRNA* beïnvloeden en daardoor wellicht bijdragen aan het ziekteproces van reuma.

Om verdere betrokkenheid van niet-coderende genen en auto-immuunziekten te bepalen zijn in **hoofdstuk 6** van dit proefschrift de expressieniveaus van alle genen in huidbiopten van patiënten met systemische sclerose onderzocht. De expressieniveaus van zo'n 60.000 genen zijn gemeten en er is met name bepaald welke niet-coderende genen anders tot expressie komen in de huidbiopten van patiënten. Voor het eerst is onderzocht welke niet-coderende genen anders zijn in patiënten en hoe deze genen mogelijk betrokken kunnen zijn bij het ziekteproces. Wij hebben aangetoond dat coderende en niet-coderende genen anders tot expressie komen in patiënten. In dit hoofdstuk komt naar voren dat verschillende antisense-genen verhoogd of verlaagd tot expressie komen in patiënten gepaard met gedereguleerde coderende sense genen, dat zijn genen die in de buurt van een antisense gen tot expressie komen. Een deel van de betrokkenheid van deze niet-coderende genen werd bevestigd in een onafhankelijke studie met Amerikaanse patiënten.

Tot slot is in dit proefschrift onderzocht hoe niet-coderende genen functioneren. Niet-coderende genen zijn niet alleen betrokken bij het ontstaan van ziekten maar ook bij allerlei belangrijke processen in ons lichaam en de ontwikkeling van een organisme. Stamcellen spelen een belangrijke rol bij ontwikkeling van organismes omdat deze cellen in staat zijn naar verschillende celtypes te veranderen. Verschillende genen zijn betrokken bij de instandhouding van stamcellen en de verandering naar andere celtypes, waaronder *Sox2*. Het *Sox2* gen wordt overlapt door een niet-coderend gen *Sox2ot* en in **hoofdstuk 7** is het mechanisme onderzocht hoe *Sox2ot* het coderende *Sox2* gen kan beïnvloeden. Wij hebben laten zien dat de expressie van *Sox2ot* negatief gecorreleerd is met de expressie Sox2. Het mechanisme dat dit bepaald lijkt afhankelijk te zijn van de enhancer-activiteit van het niet-coderende *Sox2ot*.

Concluderend, in dit proefschrift laten we zien dat naast coderende genen ook niet-coderende genen een rol spelen in auto-immuunziekten. We verwachten dat in de nabije toekomst de functies die niet-coderende genen uitoefenen verder ontrafeld worden en dat hun betrokkenheid in ziekten beter in kaart wordt gebracht. Daarnaast zien we de potentie van niet-coderende genen om te

fungeren als biomarkers en als targetgenen van toekomstige medicijnen. Zulke medicijnen kunnen dus niet-coderende genen onderscheppen of herstellen waardoor het ontstaan van ziekten kan worden uitgesteld of zelfs kan worden uitgesloten.

List of publications

1. Messemaker TC, van Leeuwen SM, van den Berg PR, 't Jong AEJ, Palstra RJ, Hoeben RC, Semrau S, Mikkers HMM. **Allele-specific repression of Sox2 through the long non-coding RNA Sox2ot.** Sci Rep. 2018 Jan 10;8(1):386. doi: 10.1038/s41598-017-18649-4.

2. Messemaker TC, Chadli L, Cai G, Goelela VS, Boonstra M, Dorjée AL, Andersen SN, Mikkers HMM, van 't Hof P, Mei H, Distler O, Draisma HHM, Johnson ME, Orzechowski NM, Simms RW, Toes REM, Aarbiou J, Huizinga TW, Whitfield ML, DeGroot J, de Vries-Bouwstra J, Kurreeman F. **Antisense long non-coding RNAs are deregulated in skin tissue of patients with systemic sclerosis.** J Invest Dermatol. 2017 Nov 24. pii: S0022-202X(17)33169-X. doi: 10.1016/j.jid.2017.09.053.

3. Messemaker TC, Mikkers HM, Huizinga TW, Toes RE, van der Helm- van Mil AH, Kurreeman F. **Inflammatory genes TNFα and IL6 display no signs of increased H3K4me3 in circulating monocytes from untreated rheumatoid arthritis patients.** Genes Immun. 2017 Sep;18(3):191-196. doi: 10.1038/gene.2017.20.

4. Messemaker TC, Frank-Bertoncelj M, Marques RB, Adriaans A, Bakker AM, Daha N, Gay S, Huizinga TW, Toes RE, Mikkers HM, Kurreeman F. **A novel long non-coding RNA in the rheumatoid arthritis risk locus TRAF1-C5 influences C5 mRNA levels.** Genes Immun. 2016 Mar;17(2):85-92. doi: 10.1038/gene.2015.54.

5. Messemaker TC, Huizinga TW, Kurreeman F. **Immunogenetics of rheumatoid arthritis: Understanding functional implications.** J Autoimmun. 2015 Nov;64:74-81. doi: 10.1016/j.jaut.2015.07.007.

6. Messemaker T, Toes RE, Mikkers HM, Kurreeman F. **Comment on "Functional Analysis of a Complement Polymorphism (rs17611) Associated with Rheumatoid Arthritis".** J Immunol. 2015 Jul 1;195(1):3-4. doi: 10.4049/jimmunol.1500822.

7. Aten E, Sun Y, Almomani R, Santen GW, Messemaker T, Maas SM, Breuning MH, den Dunnen JT. **Exome sequencing identifies a branch point variant in Aarskog-Scott syndrome.** Hum Mutat. 2013 Mar;34(3):430-4. doi: 10.1002/humu.22252.

8. van Putten M, Hulsker M, Nadarajah VD, van Heiningen SH, van Huizen E, van Iterson M, Admiraal P, Messemaker T, den Dunnen JT, 't Hoen PA, Aartsma-Rus A. **The effects of low levels of dystrophin on mouse muscle function and pathology.** PLoS One. 2012;7(2):e31937. doi: 10.1371/journal.pone.0031937.

Curriculum Vitae

Tobias Casper Messemaker werd geboren op 2 april 1988 in Katwijk aan zee. In 2006 behaalde hij zijn HAVO diploma op het Northgo college in Noordwijk en vervolgde hierna zijn studieloopbaan aan de hogeschool Leiden. Hier volgde hij de opleiding hoger laboratoriumonderwijs met moleculaire biologie als specialisatie. Hij sloot deze opleiding af in 2010 met een stage op de afdeling humane genetica van het Leids Universitair Medisch Centrum waar hij zeldzame genetische aandoeningen onderzocht in de groep van Johan den Dunnen onder leiding van Emmelien Aten.

Na het behalen van zijn bacheloropleiding vervolgde Tobias zijn studieloopbaan met de masteropleiding Biomolecular Science aan de Vrije Universiteit te Amsterdam. Tijdens deze opleiding is zijn biotechnologische kennis verbreed met een stage bij ProteoNic onder begeleiding van Raymond Verhaert. Ook onderzocht hij pluripotentheid van stamcellen tijdens een stage op de afdeling Anatomie en Embryologie in de groep van Susana Chuva de Sousa Lopes. Tobias studeerde *cum laude* af voor zijn master Biomolecular Science in 2012.

In 2013 is Tobias begonnen aan zijn promotieonderzoek aan de afdeling reumatologie en de afdeling moleculaire cel biologie van het Leids Universitair Medisch Centrum. Dit promotieproject werd in 2017 afgerond onder begeleiding van dr. Fina Kurreeman, dr. Harald Mikkers en prof. dr. Rene Toes. Momenteel is Tobias werkzaam als junior docent biomedische wetenschappen aan de Vrije Universiteit van Amsterdam.

Dankwoord

Allereerst wil ik alle patiënten bedanken die deel hebben genomen aan de studies beschreven in dit proefschrift. Ook ben ik dankbaar voor de gezonde deelnemers die vrijwillig hebben deelgenomen en het zo mogelijk hebben gemaakt om wetenschappelijk onderzoek te doen. Daarnaast is het onmogelijk een promotie te voltooien zonder daarbij hulp te aanvaarden. Om deze reden wil ik van de gelegenheid gebruik maken om de daarvoor genoemden te bedanken. Ik wil beginnen met het bedanken van mijn ouders Hennie en Margareth die altijd voor mij klaar hebben gestaan. In het bijzonder wil ik mijn vader Hennie bedanken voor zijn artistieke bijdrage aan de buitenkant van dit proefschrift, jouw bijdrage heeft extra kleur gegeven aan dit proefschrift. Ook mijn zussen Susanne en Annemarie, en mijn broer Rick, bedankt voor jullie interesse in de voortgang van mijn promotie en dat jullie er altijd in slaagden mijn gedachten over werk af te leiden, wanneer dat nodig was. Mijn vriendin Daniëlle wil ik graag in het bijzonder bedanken. Ondanks de drukke periode ben ik blij dat ik jou niet zomaar voorbij ben gelopen. Ik heb in de afgelopen jaren veel steun van jou ontvangen en veel plezier met jou gedeeld. Je zorgde ervoor dat mijn batterij weer werd opgeladen op de momenten dat dat nodig was en dat ik rust nam op de momenten dat ik je zag.

Tijdens mijn promotie ben ik begeleid door een gemotiveerd team onder begeleiding van Tom Huizinga, Rene Toes, Fina Kurreeman en Harald Mikkers en daar wil ik hen graag voor bedanken. Tom en Rene, bedankt voor jullie input tijdens de genetica-besprekingen en jullie visie, aanbevelingen en kennis om mij op koers te houden. Fina, bedankt dat ik zo ontzettend veel van jou heb kunnen leren en dat je mij enthousiast hebt gemaakt voor het bestuderen van genetica. Harald, het was mij een genoegen om met jou op het kantoor mijn promotieonderzoek te hebben volbracht. Jij was altijd bereikbaar voor vragen en input op zowel wetenschappelijk gebied als daarbuiten en zowel binnen kantooruren als daarbuiten.

Verder zijn er veel mensen van de beide afdelingen waarop ik heb gewerkt die ik nog wil bedanken. Mijn collega's van zowel de afdeling moleculaire cel biologie als de afdeling reumatologie, bedankt voor jullie input tijdens de wekelijkse, maandelijkse en jaarlijkse besprekingen. Joris, bedankt dat we altijd onze

moleculaire technieken en methodes met elkaar konden uitwisselen en bespreken. Aleida en Rosalie, bedankt voor jullie gezelschap bij celkweken en het bespreken van de nieuwste films en series. Hilde, bedankt voor de goede gesprekken over onze promotietrajecten, besprekingen en borrels en voor je hulp bij mijn zoektocht naar een nieuwe baan. Ook wil ik mijn kamergenoten Anna, Arnaud, Harald, Marjolein en Steve bedanken waarmee ik van alles kon bespreken en voor het verzorgen van een goede werksfeer. Willem en Joop bedankt voor jullie ICT-hulp en de gezellige dagjes uit. Daarnaast wil ik ook mijn andere borrelgenoten bedanken die tegelijk ook mijn teamleden waren van het MCB voetbalteam waarmee we jaarlijks werden uitgeschakeld door de beveiligers. Edwin, Reinier, Aleksandra, Sabine, Frauke, Vera, Amiet, David, Joost en Timo, bedankt.

Graag wil ik ook deze ruimte gebruiken om Marjolein en Vera in het bijzonder te noemen. Ik heb jullie voor het eerst ontmoet 2 weken voor mijn promotieonderzoek en heb sindsdien een goede band met jullie opgebouwd die mij heeft geholpen binnen en buiten mijn onderzoek. Ontzettend bedankt voor het plezier dat jullie hebben bezorgd en ik ben dankbaar dat ik jullie als paranimfen achter mij heb staan.

Tot slot, mijn beste vrienden Leithe, Michael, Arend en David. Ontzettend bedankt voor jullie vriendschap tijdens mijn promotie, ik had niet zonder jullie gekund. Arend, bedankt voor je inzet bij het *TRAF1-C5* onderzoek en alle steun en gezelligheid die je hebt gegeven naderhand. Squad hotel (Michael en Leithe) bedankt voor alle support, cover, backup en revives die ik van jullie heb ontvangen afgelopen jaren. Tot slot, Leithe, vooral jij bent van onschatbare waarde geweest voor mijn studie en promotieonderzoek doordat je mij wetenschappelijk heb gemaakt wie ik nu ben. Je hebt mij tijdens onze studie zeer gemotiveerd voor de biomedische wereld en de wetenschap en daar ben ik je voor altijd dankbaar voor.